Serial No. 10/580,979
Confirmation No. 9290
Art Unit: 1648

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
16 June 2005 (16.06.2005)

PCT

(10) International Publication Number
WO 2005/053516 A2

(51) International Patent Classification[7]: A61B

(21) International Application Number: PCT/US2004/040120

(22) International Filing Date: 1 December 2004 (01.12.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/525,989 1 December 2003 (01.12.2003) US

(71) Applicant *(for all designated States except US)*: **BOARDS OF REGENTS, THE UNIVERSITY OF TEXAS SYSTEM** [US/US]; 201 West 7th Street, Austin, TX 78701 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **LEMON, Stanley, M.** [US/US]; 1517 Bayou Shore Drive, Galveston, TX 77551 (US). **YI, MinKyung** [KR/US]; 7700 Seawall Blvd.#301, Galveston, TX 77551 (US).

(74) Agent: **PROVENCE, David, L.**; Mueting, Raasch & Gebhardt, P.A., P.O. Box 581415, Minneapolis, MN 55454-1415 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2005/053516 A2

(54) **Title:** REPLICATION COMPETENT HEPATITIS C VIRUS AND METHODS OF USE

(57) **Abstract:** The present invention provides replication competent polynucleotides that include a coding sequence encoding a hepatitis C virus polyprotein having adaptive mutations. The invention also includes methods for making replication competent polynucleotides, identifying a compound that inhibits replication of a replication competent polynucleotide, selecting a replication competent polynucleotide, and detecting a replication competent polynucleotide.

REPLICATION COMPETENT HEPATITIS C VIRUS AND METHODS OF USE

CONTINUING APPLICATION DATA

This application claims the benefit of U.S. Provisional Application Serial No. 60/525,989, filed December 1, 2003, which is incorporated by reference herein.

GOVERNMENT FUNDING

The present invention was made with government support under Grant Nos. U19-AI40035 and N01-AI25488, awarded by the National Institute of Allergy and Infectious Diseases. The Government has certain rights in this invention.

BACKGROUND

Hepatitis C virus is the most common cause of chronic viral hepatitis within the United States, infecting approximately 4 million Americans and responsible for the deaths of 8,000-10,000 persons annually due to progressive hepatic fibrosis leading to cirrhosis and/or the development of hepatocellular carcinoma. Hepatitis C virus is a single stranded, positive-sense RNA virus with a genome length of approximately 9.6 kb. It is currently classified within a separate genus of the flavivirus family, the genus *Hepacivirus*. The epatitis C virus genome contains a single large open reading frame (ORF) that follows a 5' non-translated RNA of approximately 342 bases containing an internal ribosome entry segment (IRES) directing cap-independent initiation of viral translation. The large ORF encodes a polyprotein which undergoes post-translational cleavage, under control of cellular and viral proteinases. This yields a series of structural proteins which include a core or nucleocapsid protein, two envelope glycoproteins, E1 and E2, and at least six nonstructural replicative proteins. These include NS2 (which with the adjacent NS3 sequence demonstrates *cis*-active metalloproteinase activity at the NS2/NS3 cleavage site), NS3 (a serine proteinase/NTPase/RNA helicase), NS4A (serine proteinase accessory factor), NS4B, NS5A, and NS5B (RNA-dependent RNA polymerase).

With the exception of the 5' non-translated RNA, there is substantial genetic heterogeneity among different stains of hepatitis C virus. Phylogenetic analyses have led to the classification of hepatitis C virus strains into a series of genetically distinct "genotypes," each of which contains a group of genetically related viruses. The genetic distance between some of these genotypes is large enough to suggest that there may be biologically significant serotypic differences as well. There is little understanding of the extent to which infection with a virus of any one genotype might confer protection against viruses of a different genotype.

The currently available therapy of interferon in combination with ribavirin has poor response rate against most prevalent strains of HCV, genotype 1a and 1b. Establishment of selectable subgenomic replicon systems has advanced the study of HCV RNA replication. However, only replicons of genotype 1b strains are readily available, and extension of replicon systems to other genotypes has been largely unsuccessful. Considering the nature of high genetic variability of HCV, HCV replication systems derived from other genotypes will be very helpful in the effort of drug discovery. In support with this notion, chimeric replicons containing a genotype 1a polymerase in the background of a genotype 1b replicon were more resistant to interferon treatment in vitro than the replicon derived from a genotype 1b HCV. Extension of replicon system to other genotypes are also necessary to understand the mechanism of HCV RNA replication and the contribution of variable sequences in that process.

Recently two groups reported the generation of genotype 1a replication system using highly permissive sublines of Huh-7 cells. Blight et al. (J. Virol. 77, 3181-3190 (2003)) were able to select G418 resistant colonies supporting replication of genotype 1a derived subgenomic replicons in a hyper-permissive Huh7 subline, Huh-7.5, that was generated by curing an established G418-resistant replicon cell line of the cubgenomic Con1 replicon RNA that had been used to select it by treatment with interferon-alpha (Blight et al., J. Virol., 76, 13001-13014 (2002)). Sequence analysis of replicating HCV RNAs inside of such selected cell lines showed that the most common critical mutations were located at amino acid position 470 of NS3 (P1496L) within domain II of the NS3 helicase, and the NS5A mutation (S2204I). In other case, Grobler et al. (J. Biol. Chem., 278,16741-16746 (Feb, 2003)), used a systematic mutational approach to reach

the similar conclusion that both P1496L and S2204I combination was necessary to get genotype 1a replication in a highly permissive Huh-7 subline which was selected in an independent but similar way. However, genotype1a RNAs with these two enhanced mutations does not undergo replication in the Huh-7 cell line, indicating limited usefulness of this system.

SUMMARY

The present invention provides replication competent polynucleotides. The replication competent polynucleotides include a 5' non-translated region (NTR), a 3' NTR, and a first coding sequence present between the 5' NTR and 3' NTR and encoding a hepatitis C virus polyprotein. The 5' NTR, the 3' NTR, and the nucleotide sequence encoding the polyprotein may be genotype 1a. The polyprotein includes an isoleucine at about amino acid 2204, and further includes an adaptive mutation. The adaptive mutation can be an arginine at about amino acid 1067, an arginine at about amino acid 1691, a valine at about amino acid 2080, an isoleucine at about amino acid 1655, an arginine at about amino acid 2040, an arginine at about amino acid 1188, or a combination thereof. The polyprotein may be a subgenomic polyprotein. The polyprotein may include the cleavage products core, E1, E2, P7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B. The replication competent polynucleotides may further include a second coding sequence. The second coding sequence can encode, for instance, a marker or a transactivator. The replication competent polynucleotides may further include a nucleotide sequence having cis-acting ribozyme activity, wherein the nucleotide sequence is located 3' of the 3' NTR.

Also provided by the present invention are methods for making a replication competent polynucleotide, and the resulting replication competent polynucleotide. The methods include providing a polynucleotide having a 5' NTR, 3' NTR, a first coding sequence present between the 5' NTR and 3' NTR and encoding a hepatitis C virus polyprotein. Typically, the 5' NTR, polyprotein, and 3' NTR are genotype 1a. The polyprotein includes a serine at about amino acid 2204, a glutamine at about amino acid 1067, a lysine at about amino acid 1691, a phenylalanine at about amino acid 2080, a valine at about amino acid 1655, a lysine at about amino acid 2040, or a glycine at about amino acid 1188. The

method also includes altering the coding sequence such that the polyprotein encoded thereby includes an isoleucine at amino acid 2204, and an adaptive mutation. The polyprotein may be a subgenomic polyprotein. The polyprotein may include the cleavage products core, E1, E2, P7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B.

The present invention further provides methods for identifying a compound that inhibits replication of a replication competent polynucleotide. The method includes contacting a cell containing a replication competent polynucleotide with a compound, incubating the cell under conditions wherein the replication competent polynucleotide replicates in the absence of the compound, and detecting the replication competent polynucleotide, wherein a decrease of the replication competent HCV polynucleotide in the cell contacted with the compound compared to the replication competent polynucleotide in a cell not contacted with the compound indicates the compound inhibits replication of the replication competent polynucleotide. The detecting of the replication competent polynucleotide can include, for instance, nucleic acid amplification or identifying a marker encoded by the replication competent polynucleotide or by the cell containing the replication competent polynucleotide.

Also provided by the present invention are methods for selecting a replication competent polynucleotide. The method includes incubating a cell containing a polynucleotide including a 5'NTR, a 3'NTR, and a first coding sequence present between the 5'NTR and 3'NTR and encoding a hepatitis C virus polyprotein, and a second coding sequence. The polyprotein includes an isoleucine at about amino acid 2204, and further includes an adaptive mutation. The second coding sequence encodes a selectable marker conferring resistance to a selecting agent that inhibits replication of a cell that does not express the selectable marker. The method also includes detecting a cell that replicates in the presence of the selecting agent, wherein the presence of such a cell indicates the polynucleotide is replication competent. The method may further include obtaining a virus particle produced by the cell, exposing a second cell to the isolated virus particle and incubating the second cell in the presence of the selecting agent, and detecting a second cell that replicates in the presence of the selecting agent, wherein the presence of such a cell indicates the replication competent polynucleotide in the first cell produces an infectious virus particle.

The present invention also provides methods for detecting a replication competent polynucleotide, including incubating a cell containing a replication competent polynucleotide. The replication competent polynucleotide includes a 5' NTR, a 3' NTR, and a first coding sequence present between the 5' NTR and 3' NTR and encoding a hepatitis C virus polyprotein, and a second coding sequence encoding a transactivator. The cell includes a transactivated coding region and an operator sequence operably linked to the transactivated coding region, and the transactivated coding region encodes a detectable marker, wherein the transactivator alters transcription of the transactivated coding region. The method further includes detecting the detectable marker, wherein the presence of the detectable marker indicates the cell includes a replication competent polynucleotide.

Definitions

As used herein, the term "replication competent polynucleotide" refers to a polynucleotide that replicates when present in a cell. For instance, a complementary polynucleotide is synthesized. As used herein, the term "replicates *in vitro*" indicates the polynucleotide replicates in a cell that is growing in culture. The cultured cell can be one that has been selected to grow in culture, including, for instance, an immortalized or a transformed cell. Alternatively, the cultured cell can be one that has been explanted from an animal. "Replicates *in vivo*" indicates the polynucleotide replicates in a cell within the body of an animal, for instance a primate (including a chimpanzee) or a human. In some aspects of the present invention, replication in a cell can include the production of infectious viral particles, i.e., viral particles that can infect a cell and result in the production of more infectious viral particles.

As used herein, the term "polynucleotide" refers to a polymeric form of nucleotides of any length, either ribonucleotides or deoxynucleotides, and includes both double- and single-stranded DNA and RNA. A polynucleotide may include nucleotide sequences having different functions, including for instance coding sequences, and non-coding sequences such as regulatory sequences and/or non-translated regions. A polynucleotide can be obtained directly from a natural source, or can be prepared with the aid of recombinant, enzymatic, or chemical techniques. A polynucleotide can be linear or circular in topology and can be, for

example, a portion of a vector, such as an expression or cloning vector, or a fragment.

The terms "coding region" and "coding sequence" are used interchangeably and refer to a polynucleotide region that encodes a polypeptide and, when placed under the control of appropriate regulatory sequences, expresses the encoded polypeptide. The boundaries of a coding region are generally determined by a translation start codon at its 5' end and a translation stop codon at its 3' end. A coding region can encode one or more polypeptides. For instance, a coding region can encode a polypeptide that is subsequently processed into two or more polypeptides. A regulatory sequence or regulatory region is a nucleotide sequence that regulates expression of a coding region to which it is operably linked. Nonlimiting examples of regulatory sequences include promoters, transcription initiation sites, translation start sites, internal ribosome entry sites, translation stop sites, and terminators. "Operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. A regulatory sequence is "operably linked" to a coding region when it is joined in such a way that expression of the coding region is achieved under conditions compatible with the regulatory sequence.

"Polypeptide" as used herein refers to a polymer of amino acids and does not refer to a specific length of a polymer of amino acids. Thus, for example, the terms peptide, oligopeptide, protein, polyprotein, proteinase, and enzyme are included within the definition of polypeptide. This term also includes post-expression modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations and the like. A "hepatitis C virus polyprotein" refers to a polypeptide that is post-translationally cleaved to yield more than one polypeptide.

The terms "5' non-translated RNA," "5' non-translated region," "5' untranslated region" and "5' noncoding region" are used interchangeably, and are terms of art (see Bukh et al., Proc. Nat. Acad. Sci. U S A, 89, 4942-4946 (1992)). The term refers to the nucleotides that are at the 5' end of a replication competent polynucleotide.

The terms "3' non-translated RNA," "3' non-translated region," and "3' untranslated region" are used interchangeably, and are terms of art. The term

refers to the nucleotides that are at the 3' end of a replication competent polynucleotide.

Unless otherwise specified, "a," "an," "the," and "at least one" are used interchangeably and mean one or more than one.

BRIEF DESCRIPTION OF THE FIGURES

Figure. 1. Organization of the selectable subgenomic dicistronic HCV replicons, Bpp-Ntat2ANeo/SI (identical to Ntat2Aneo/SI in Yi et al., Virol., 302, 197-210 (2002)), Htat2ANeo/SI, and Bpp-Htat2ANeo/SI, in which most of the nonstructural protein-coding region and the 3'NTR are derived from the H77c HCV genotype 1a sequence. The two large ORFs are shown as rectangles, with nontranslated RNA segments shown as lines. The segment of the 3' ORF labeled 'pp' ('proximal protease') encodes the amino terminus of the NS3 protein (residues 1 to 75). 'Bpp' indicates that this region is derived from the HCV Con1 sequence. Both replicons contain the S2204I mutation in NS5A (S→I). 'δ' Indicates the hepatitis delta ribozyme sequence introduced downstream of the 3' terminus of the HCV sequence that produces an exact 3' end.

Figure. 2. Transient HCV RNA replication assay. Shown is the expression of SEAP by En5-3 cells following transfection with the chimeric 1a replicon Bpp-Htat2ANeo/SI and Bpp-Htat2ANeo/KR/SI, which carries an additional K1691R mutation in NS3 that was identified following selection of G418-resistant cells following transfection with Bpp-Htat2ANeo/SI. As controls, SEAP expression is shown following transfection of cells with the highly replication competent 1b replicon, Bpp-Ntat2ANeo/SI, and a related replication defective ΔGDD mutant; also shown in SEAP expression by normal En5-3 cells. Results shown represent the mean values obtained from triplicate cultures transfected with each RNA. SI, S2204 adaptive mutation; KR, K1691R adaptive mutation.

Figure. 3. (A) Schematic depicting the organization of the 5' end of the second ORF in subgenomic chimeric replicons containing most (Bpp-H34A-Ntat2ANeo/SI) or all (Hpp-H34A-Ntat2ANeo/SI) of the H77 genotype 1a NS34A-coding sequence in the background of the genotype 1b Bpp-

Ntat2ANeo/SI. Genotype 1a sequence (H77) is shown as an open box, genotype 1b sequence (Con1 or HCV-N) as a shaded box. 'Bpp' indicates the presence of genotype 1b sequence from the Con1 strain of HCV in the 5' proximal protease coding sequence, whereas 'Hpp' indicates that this sequence is derived from the genotype 1a H77 sequence. Approximate locations are shown for the adaptive mutations Q1067R (Q→R) and G1188R (G→R), identified in G418-resistant cell clones selected following transfecttion of Hpp-H34A-Ntat2ANeo/SI. (B) SEAP activity present in supernatant culture fluids collected at 24 hr intervals following transfection of En5-3 cells with various chimeric 1a-1b replicons including Bpp-H34A-Ntat2ANeo/SI, Hpp-H34A-Ntat2ANeo/SI, Hpp-H34A-Ntat2ANeo/QR/SI, and Hpp-H34A-Ntat2ANeo/GR/SI. Control cells were transfected with Bpp-Ntat2ANeo/SI and the replication defective ΔGDD mutant. See legend to Fig. 2 for further details. SI, S2204 adaptive mutation; QR, Q1067R adaptive mutation; and GR, G1188R adaptive mutation.

Figure 4. Impact of adaptive mutations on replication competence of the subgenomic genotype 1a replicon, Htat2ANeo/SI. (A) Location of various adaptive mutations within the second ORF (derived entirely from the genotype 1a H77sequence): Q1067R, P1496L (NS3); K1691R (NS4A); and F2080V and S2204I (NS5A). (B) Transient HCV RNA replication assay. SEAP activity in culture supernatants collected at 12-24 hr intervals following electroporation of En5-3 cells with the 1a replicon Htat2ANeo carrying the indicated combinations of the adaptive mutations shown in panel A. Cells were also transfected with genotype 1b Bpp-Ntat2ANeo/SI replicon RNA as a reference. (C) Summary of the replication phenotypes of genotype 1a replicon Htat2ANeo RNAs containing various combinations of adaptive mutations: (-) no detectable replication, (+) modest increase in SEAP expression above background days 3-5, and (+++) >10-fold increase in SEAP expression above background 7 days after transfection in the transient replication assay (see panel B). SI, S2204 adaptive mutation; QR, Q1067R adaptive mutation; PL, P1496L adaptive mutation; KR, K1691R adaptive mutation; and FV, F2080V adaptive mutation.

Figure. 5. Adaptive mutations within the polyprotein do not influence the efficiency of polyprotein translation under control of the EMCV IRES. Shown is

an SDS-PAGE gel loaded with products of in vitro translation reactions programmed with RNAs derived from Bpp-Ntat2ANeo (lane1), Bpp-Htat2ANeo (lanes 2 and 3), Htat2ANeo (lanes 4 to 8), or Bpp-Ntat2ANeo/ΔGDD (lane 9) RNAs carrying various combinations of adaptive mutations (Q1067R, K1691R, F2080V, or S2204I) as indicated. The schematic at the top of the figure indicates the location of these mutations within the polyprotein. 'pp' indicates the RNA segment encoding the amino terminal 75 residues of NS3, while 'NS' indicates the remainder of the RNA segment encoding the nonstructural proteins. H = genotype 1a H77 sequences, B = genotype 1b Con1 sequences, and N = genotype 1b HCV-N sequences. Location of NS3 and Neo product is indicated at the side of gel.

Figure 6. Impact of additional adaptive mutations on replication competence of the subgenomic genotype 1a replicon, Htat2ANeo/QR/KR/SI (see Fig. 4). (A) Location of various adaptive mutations within the second ORF (derived entirely from the genotype 1a H77sequence): Q1067R, V1655I (NS3); K1691R (NS4A); and K2040R (KR^{5A}), F2080V and S2204I (NS5A). (B) Transient HCV RNA replication assay. SEAP activity in culture supernatants collected at 12-24 hr intervals following electroporation of En5-3 cells with the 1a replicon Htat2ANeo carrying the indicated combinations of the adaptive mutations shown in panel A. Cells were also transfected with genotype 1b Bpp-Ntat2ANeo/SI replicon RNA as a reference. QR, Q1067R adaptive mutation; VI, V1655I adaptive mutation; KR, K1691R adaptive mutation; KR^{5A}, K2040 adaptive mutation; FV, F2080V adaptive mutation; and SI, S2204I adaptive mutation.

Figure 7. Northern analysis of HCV RNA abundance 4 days following transfection of normal Huh7 or En5-3 cells with the indicated dicistronic subgenomic and monocistronic genome length HCV RNAs: (lane 1), normal cells; (lane 2), the subgenomic replicon, Htat2ANeo/SI; (lanes 2-5), Htat2ANeo/SI replicon RNAs carrying the indicated combinations of mutations; (lane 6), nonreplicating Htat2ANeo/QR/VI/KR//KR5A/SI/AAG; (lanes 7) genome-length H77c RNA; (lanes 8-10), genome-length H77c RNA containing the indicated combinations of mutations; (lane 11), genome-length H77 RNA containing the lethal NS5B mutation; (lanes 12 and 13) subcontrol genomic and genome-length

synthetic RNA transcripts. Blots were probed with a genotype 1a probe derived from the NS5B coding sequence for detection of HCV-specific sequence (top panels); blots were also probed for b-actin message to assess RNA loading (lower panels). At the top of the figure is shown the En5-3 cell culture supernatant fluid SEAP activity induced by replicating subgenomic RNAs at the time of cell harvest. SI, S2204 adaptive mutation; QR, Q1067R adaptive mutation; KR, K1691R adaptive mutation; and FV, F2080V adaptive mutation.

Figure 8. Structure of the NS3/4A serine protease/helicase enzyme complex derived from the genotype 1b BK strain of HCV (PDP 1CU1), with the locations of adaptive mutations highlighted. (A) Wire diagram of structure showing the NS3 helicase domain (H) and the protease domain (P). The NS4A cofactor polypeptide (NS4A) is shown in space-filling view, with the NS3 protease active site residues (Active Site) shown in space-filling view. Adaptive mutations identified in this study (Q1067, G1188, V1655, and K1691) cluster near the protease active site or at sites involved in substrate recognition, including the mutations in the NS3 protease domain at Gln-1067, Gly-1188 and near the carboxyl terminus of NS3 in the helicase domain at Val-1655. The NS4A adaptive mutation at Lys-1691 is just beyond the surface of the protease, at the site of exit of the NS4A strand. Adaptive mutations within the NS3 helicase domain that were identified in other studies, S1222, A1226, and P1496 are shown in space-filling view, and are not close to the protease active site. (B) Space-filling view of the structure shown in panel A, in which the adaptive mutations and active site have similar shading. The NS3/4A adaptive mutations identified in this study (Q1067R, G1188R, V1655I, and K1691R) all occur at solvent accessible residues on this side of the molecule. (C) Flip-view of the structure shown in panel B, rotated approximately 180 degrees. The helicase adaptive mutations identified in previous studies are located on the surface of the helicase, distant from the protease active site. Note that in the sequence of the genotype 1b BK strain of HCV, Pro-1496 is Arg (referred to as P1496(R) in the figure, and Lys-1691 is Ser (referred to as K1691(S) in the figure).

Figure 9. Nucleotide sequence of HIVSEAP (SEQ ID NO:7). The HIV long terminal repeat (LTR) is depicted at nucleotides 1-719, and secretory alkaline phosphatase is encoded by the nucleotides 748-2239.

Figure 10. 10A, nucleotide sequence of a 3' NTR (SEQ ID NO:8); 10B, nucleotide sequence of a 5' NTR (SEQ ID NO:9).

Figure 11. 11A, nucleotide sequence of a genomic length (full length) hepatitis C virus, genotype 1a (SEQ ID NO:11); 11B, the amino acid sequence of the HCV polyprotein (SEQ ID NO:12) encoded by the coding region present in SEQ ID NO:11.

Figure 12. 12A, nucleotide sequence of Htat2ANeo (SEQ ID NO:13), where nucleotide 1-341 are the 5'NTR, nucleotides 342-1454 are the tat2ANeo (termination codon at 1455-1457), nucleotides 1458-2076 are the EMCV IRES, nucleotides 2080-8034 encode the HCV polyprotein (initiation codon at nucleotides 2077-2079 and termination codon at nucleotides 8035-8037), nucleotides 8038-8259 are the 3'NTR, and nucleotides 8260-8345 are the HDV delta ribozyme (plasmid vector sequences are shown at nucleotides 8346-11240); 12B, the amino acid sequence of the HCV polyprotein (SEQ ID NO:14) encoded by the coding region present in SEQ ID NO:13.

Figure 13. Nucleotide (SEQ ID NO:1) of Hepatitis C virus strain H77 and amino acid sequence (SEQ ID NO:2) encoded by nucleotides 342 - 9377.

Figure 14. Nucleotide (SEQ ID NO:3) of Hepatitis C virus strain H and amino acid sequence (SEQ ID NO:4) encoded by nucleotides 342 - 9377.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS OF THE INVENTION

The present invention provides replication competent polynucleotides. The polynucleotides include a 5' non-translated region (NTR), a 3' NTR, and a coding sequence present between the 5' NTR and 3' NTR. The replication

competent polynucleotides of the present invention are based on hepatitis C virus (HCV), a positive-strand virus. While the ability of a polynucleotide to replicate typically requires the presence of the positive-strand RNA polynucleotide in a cell, it is understood that the term "replication competent polynucleotide" also includes the complement thereof (i.e., the negative-sense RNA), and the corresponding DNA sequences of the positive-sense and the negative-sense RNA sequences. Optionally, a replication competent polynucleotide may be isolated. "Isolated" means a biological material, for instance a polynucleotide, polypeptide, or virus particle, that has been removed from its natural environment. For instance, a virus that has been removed from an animal or from cultured cells in which the virus was propagated is an isolated virus. An isolated polypeptide or polynucleotide means a polypeptide or polynucleotide that has been either removed from its natural environment, produced using recombinant techniques, or chemically or enzymatically synthesized. A "purified" biological material is one that is at least 60% free, preferably 75% free, and most preferably 90% free from other components with which it is naturally associated.

The coding sequence encodes a hepatitis C virus polyprotein. In some aspects of the invention, the HCV polyprotein can yield the following polypeptides; core (also referred to as C or nucleocapsid), E1, E2, P7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B. Optionally, a full length HCV polyprotein also yields protein F (see Xu et al., *EMBO J.*, 20, 3840-3848 (2001). In some aspects of the present invention, an HCV polyprotein is shortened and yields a subset of polypeptides, and typically does not include polypeptides encoded by the amino terminal end of the full length HCV polyprotein. Thus, a hepatitis C virus polyprotein may encode the polypeptides E1, E2, P7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B; E2, P7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B; P7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B; NS2, NS3, NS4A, NS4B, NS5A, and NS5B; or NS3, NS4A, NS4B, NS5A, and NS5B. The hepatitis C virus encoding such a shortened HCV polyprotein may be referred to as a subgenomic hepatitis C virus, and the shortened HCV polyprotein may be referred to as a subgenomic HCV polyprotein. In other aspects of the invention, a replication competent polynucleotide encodes an HCV polyprotein that does not include polypeptides present in an internal portion of a hepatitis C virus polyprotein. Thus, a

subgenomic hepatitis C virus polyprotein may encode, for instance, the polypeptides NS3, NS4A, NS4B, and NS5B.

In those aspects of the invention where the replication competent polynucleotide includes a coding region that encodes less than a full length HCV polyprotein, the 5' end of the coding region encoding the HCV polyprotein may further include about 33 to about 51 nucleotides, or about 36 to about 48 nucleotides, that encode the first about 11 to about 17, or about 12 to about 16, amino acids of the core polypeptide. The result is a fusion polypeptide made up of amino terminal amino acids of the core polypeptide and the first polypeptide encoded by the first cleavage product of the polyprotein, e.g., E1, or E2, or P7, or NS2, etc.

A polyprotein that can yield the core, E1, E2, P7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B polypeptides (a full length polyprotein) is typically between about 3000 and 3033 amino acids in length, preferably about 3011 amino acids in length. The relationship between such a polyprotein and the corresponding residues of the individual polypeptides resulting after post-translational processing is shown in Table 1. This numbering system is used herein when referring to a full length polyprotein, and when referring to a polyprotein that contains a portion of the full length polyprotein. For instance, in those aspects of the invention where the replication competent polynucleotide includes a coding sequence encoding an HCV polyprotein that yields the cleavage products NS3, NS4A, NS4B, NS5A, and NS5B and there is no fusion polypeptide made up of amino terminal amino acids of the core polypeptide and the cleavage product NS3, the first amino acid of the NS3 polypeptide is considered to be about residue number 1027. A person of ordinary skill in the art recognizes that this numbering system can vary between members of different genotypes, and between members of the same genotype, thus the numbers shown in Table 1 are approximate, and can vary by 1, 2, 3, 4, or about 5.

Table 1. Correspondence between amino acids of polyprotein and individual polypeptides after processing.

Amino acids of HCV polyprotein[a]	Corresponding polypeptide after processing
1-191	Core
192-383	E1
384-746	E2
747-809	P7
810-1026	NS2
1027-1657	NS3
1658-1711	NS4A
1712-1972	NS4B
1973-2420	NS5A
2421-3011	NS5B

[a] Refers to the approximate amino acid number prior to cleavage of the polyprotein where the first amino acid is the first amino acid of the polyprotein expressed by the HCV at Genbank Accession number AF011751 and Genbank Accession number M67463.

A replication competent polynucleotide of the present invention includes at least one adaptive mutation. As used herein, an adaptive mutation is a change in the amino acid sequence of the polyprotein that increases the ability of a replication competent polynucleotide to replicate compared to a replication competent polynucleotide that does not have the adaptive mutation. One adaptive mutation that a replication competent polynucleotide of the present invention typically includes is an isoleucine at about amino acid 2204, which is about amino acid 232 of NS5A. Most clinical HCV isolates and molecularly cloned laboratory HCV strains include a serine at this position, and this mutation has been referred to in the art as S2204I. In most replication competent polynucleotides, the location of this adaptive mutation can also be determined by locating the amino acid sequence SSSA beginning at about amino acid 2200 in the HCV polyprotein, where the amino acid immediately following the SSSA sequence is isoleucine.

A replication competent polynucleotide of the present invention may also include one or more of the adaptive mutations described herein, or a combination thereof. The first such adaptive mutation is an arginine at about amino acid 1067, which is about amino acid 41 of NS3. Most clinical HCV isolates and molecularly cloned laboratory HCV strains include a glutamine at this position, thus this mutation can be referred to as Q1067R. In most replication competent polynucleotides, the location of this adaptive mutation can also be determined by locating the amino acid sequence STAT beginning at about amino acid 1063 in

the HCV polyprotein, where the amino acid immediately following the STAT sequence is arginine. The second adaptive mutation is an arginine at about amino acid 1691, which is about amino acid 34 of NS4A. Most clinical HCV isolates and molecularly cloned laboratory HCV strains include a lysine at this position, thus this mutation can be referred to as K1691R. In most replication competent polynucleotides, the location of this adaptive mutation can also be determined by locating the amino acid sequence VLSG beginning at about amino acid 1687 in the HCV polyprotein, where the amino acid immediately following the VLSG sequence is arginine. The third adaptive mutation is a valine at about amino acid 2080, which is about amino acid 108 of NS5A. Most clinical HCV isolates and molecularly cloned laboratory HCV strains include a phenylalanine at this position, thus this mutation can be referred to as F2080V. In most replication competent polynucleotides, the location of this adaptive mutation can also be determined by locating the amino acid sequence ALWR beginning at about amino acid 2081 in the HCV polyprotein, where the amino acid immediately before the ALWR sequence is valine. A fourth adaptive mutation is an isoleucine at about amino acid 1655, which is about amino acid 629 of NS3. Most clinical HCV isolates and molecularly cloned laboratory HCV strains include a valine at this position, thus this mutation can be referred to as V1655I. In most replication competent polynucleotides, the location of this adaptive mutation can also be determined by locating the amino acid sequence ADLE beginning at about amino acid 2051 in the HCV polyprotein, where the amino acid immediately after the ADLE sequence is isoleucine. A fifth adaptive mutation is an arginine at about amino acid 2040, which is about amino acid 68 of NS5A. Most clinical HCV isolates and molecularly cloned laboratory HCV strains include a lysine at this position, thus this mutation can be referred to as K2040R. In most replication competent polynucleotides, the location of this adaptive mutation can also be determined by locating the amino acid sequence GHVXN beginning at about amino acid 2037 in the HCV polyprotein, where the X in the amino acid is arginine. A sixth adaptive mutation is an arginine at about amino acid 1188, which is about amino acid 162 of NS3. Most clinical HCV isolates and molecularly cloned laboratory HCV strains include a glycine at this position, thus this mutation can be referred to as G1188R. In most replication competent polynucleotides, the location of this adaptive mutation can also be determined by

locating the amino acid sequence VCTR beginning at about amino acid 1184 in the HCV polyprotein. In some aspects, the replication competent polynucleotide of the present invention includes the Q1067R and K1691R adaptive mutations, as well as the S2204I adaptive mutation. These adaptive mutations are summarized in Table 2. A person of ordinary skill in the art recognizes that the precise location of these cell culture adaptive mutations can vary between members of different genotypes, and between members of the same genotype, thus the numbers shown in Table 2 are approximate, and can vary by 1, 2, 3, 4, or about 5.

Table 2. Adaptive Mutations

Symbol[1]	Protein / Residue[2]	Mutation[3]
QR	NS3/41	Q1067R
GR	NS3/162	G1188R
VI	NS3/629	V1655I
KR	NS4A/34	K1691R
KR[5A]	NS5A/68	K2040R
FV	NS5A/108	F2080V
SI	NS5A/232	S2204I

[1]Symbol used to designate presence in RNA transcripts.

[2]Residue refers to position in protein after post-translational cleavage of the H77c polyprotein (GenBank accession AF011751).

[3]Number refers to position of mutation in H77c polyprotein before post-translational cleavage (GenBank accession AF011751).

There are many other adaptive mutations known to the art, and the replication competent polynucleotides of the present invention may include one or more of those adaptive mutations. Examples of known adaptive mutations can be found in, for instance, Bartenschlager (U.S. Patent 6,630,343), Blight et al. (Science, 290, 1972-1975 (2000)), Lohmann et al., (Abstract P038, 7th International Meeting on Hepatitis C virus and Related viruses (Molecular Virology and Pathogenesis), December 3-7 (2000)), Guo et al. (Abstract P045, 7th International Meeting on Hepatitis C virus and Related viruses (Molecular

Virology and Pathogenesis), December 3-7 (2000)), Blight et al., (J. Virol. 77, 3181-3190 (2003)), Gu et al., (J. Virol. 77, 5352-5359 (2003)), and Grobler et al. ,(J. Biol. Chem., 278,16741-16746 (Feb, 2003).

It is expected that polynucleotides encoding an HCV polyprotein can be obtained from different sources, including molecularly cloned laboratory strains, for instance cDNA clones of HCV, and clinical isolates. Examples of molecularly cloned laboratory strains include the HCV that is encoded by pCV-H77C (Yanagi et al., *Proc. Natl. Acad. Sci. USA*, 94, 8738-8743 (1997), Genbank accession number AF011751, SEQ ID NO:1), and pHCV-H (Inchauspe et al., *Proc. Natl. Acad. Sci. USA*, 88, 10292-10296 (1991), Genbank accession number M67463, SEQ ID NO:3). Clinical isolates can be from a source of infectious HCV, including tissue samples, for instance from blood, plasma, serum, liver biopsy, or leukocytes, from an infected animal, including a human or a primate. It is also expected that the polynucleotide encoding the HCV polyprotein present in a replication competent polynucleotide can be prepared by recombinant, enzymatic, or chemical techniques. The nucleotide sequence of molecularly cloned laboratory strains and clinical isolates can be modified to encode an HCV polyprotein that includes the S2204I adaptive mutation and one or more of the adaptive mutations described herein. Such methods are routine and known to the art and include, for instance, PCR mutagenesis.

The present invention further includes replication competent polynucleotides encoding an HCV polyprotein having similarity with the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4 (in the case of a full length polyprotein), or a portion thereof (in the case of an HCV polyprotein encoding, for instance, NS3, NS4A, NS4B, NS5A, and NS5B, and not encoding core, E1, E2, P7, and NS2). The similarity is referred to as structural similarity and is generally determined by aligning the residues of the two amino acid sequences (i.e., a candidate amino acid sequence and the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, or a portion thereof) to optimize the number of identical amino acids along the lengths of their sequences; gaps in either or both sequences are permitted in making the alignment in order to optimize the number of identical amino acids, although the amino acids in each sequence must nonetheless remain in their proper order. A candidate amino acid sequence is the amino acid sequence being compared to an amino acid sequence present in SEQ ID NO:2,

SEQ ID NO:4, or a portion thereof. A candidate amino acid sequence can be isolated from a cell infected with a hepatitis C virus, or can be produced using recombinant techniques, or chemically or enzymatically synthesized. Preferably, two amino acid sequences are compared using the Blastp program of the BLAST 2 search algorithm, as described by Tatusova, et al. (*FEMS Microbiol Lett* 1999, 174:247-250), and available at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. Preferably, the default values for all BLAST 2 search parameters are used, including matrix = BLOSUM62; open gap penalty = 11, extension gap penalty = 1, gap x_dropoff = 50, expect = 10, wordsize = 3, and optionally, filter on. In the comparison of two amino acid sequences using the BLAST search algorithm, structural similarity is referred to as "identities." An HCV polyprotein may include an amino acid sequence having a structural similarity with SEQ ID NO:2, SEQ ID NO:4, or a portion thereof, of at least about 90 %, for example 91%, 92%, 93% identity, and so on to 100 % identity. A replication competent polynucleotide having a 5' NTR of SEQ ID NO:9, a 3' NTR of SEQ ID NO:8, and HCV polyprotein with structural similarity with SEQ ID NO:2, SEQ ID NO:4, or a portion thereof, is replication competent in a cell derived from a human hepatoma such as Huh-7 and Huh-7.5. An HCV polyprotein having structural similarity with the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, or a portion thereof, includes the S2204I adaptive mutation and one or more of the adaptive mutations described herein. Such an HCV polyprotein may optionally include other adaptive mutations.

In some aspects, the coding sequence of a replication competent polynucleotide of the present invention that encodes a hepatitis C virus polyprotein is not a specific genotype. For instance, a polynucleotide encoding an HCV polyprotein present in a replication competent polynucleotide of the present invention can be genotype 1a, 1b, 1c, 2a, 2b, 2c, 3a, 3b, 4, 5a, or 6a (as defined by Simmonds, *Hepatology*, 21, 570-583 (1995)). In other aspects, the HCV polyprotein is genotype 1a. Methods for determining the genotype of a hepatitis C virus are routine and known to the art and include, for instance, serotyping the virus particle using antibody, and/or evaluation of the nucleotide sequence by, for instance, polymerase chain reaction assays (see Simmonds, *J. Hepatol.*, 31(Suppl. 1), 54-60 (1999)).

The present invention includes polynucleotides encoding an amino acid sequence having similarity to an HCV polyprotein. The similarity is referred to as structural similarity and is determined by aligning the residues of two polynucleotides (e.g., the nucleotide sequence of the candidate coding region and nucleotides 342 - 9377 of SEQ ID NO:1 or nucleotides 342 - 9377 of SEQ ID NO:3) to optimize the number of identical nucleotides along the lengths of their sequences; gaps in either or both sequences are permitted in making the alignment in order to optimize the number of shared nucleotides, although the nucleotides in each sequence must nonetheless remain in their proper order. A candidate coding region is the coding region being compared to a coding region present in SEQ ID NO:1 (e.g., nucleotides 342 - 9377 of SEQ ID NO:1). A candidate nucleotide sequence can be isolated from a cell, or can be produced using recombinant techniques, or chemically or enzymatically synthesized. Preferably, two nucleotide sequences are compared using the Blastn program of the BLAST 2 search algorithm, as described by Tatusova, et al. (*FEMS Microbiol Lett* 1999, 174:247-250), and available at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. Preferably, the default values for all BLAST 2 search parameters are used, including reward for match = 1, penalty for mismatch = -2, open gap penalty = 5, extension gap penalty = 2, gap x_dropoff = 50, expect = 10, wordsize = 11, and optionally, filter on. In the comparison of two nucleotide sequences using the BLAST search algorithm, structural similarity is referred to as "identities."

The present invention also includes polynucleotides encoding the HCV polyproteins described herein, including, for instance, the polyproteins having the amino acid sequence shown in SEQ ID NO:2 and SEQ ID NO:4. An example of the class of nucleotide sequences encoding each of these polyproteins are nucleotides 342 - 9377 of SEQ ID NO:1 and nucleotides 342 - 9377 of SEQ ID NO:3, respectively. These classes of nucleotide sequences are large but finite, and the nucleotide sequence of each member of the class can be readily determined by one skilled in the art by reference to the standard genetic code.

A replication competent polynucleotide of the present invention includes a 5' non-translated region (NTR) (see Smith et al., *J. Gen. Virol.*, 76, 1749-1761 (1995)). A 5' NTR is typically about 341 nucleotides in length. A replication competent polynucleotide of the present invention also includes a 3' NTR. A 3' NTR typically includes, from 5' to 3', nucleotides of variable length and sequence

(referred to as the variable region), a poly-pyrimidine tract (the poly U-UC region), and a highly conserved sequence of about 100 nucleotides (the conserved region) (see, for instance, Lemon et al., U.S. Published Application US 2003 0125541, and Yi and Lemon, J. Virol., 77, 3557-3568 (2003)). The variable region begins at about the first nucleotide following the stop codon of the HCV polyprotein, and generally ends immediately before the nucleotides of the poly U-UC region. The poly U-UC region is a stretch of predominantly U residues, CU residues, or $C(U)_n$-repeats. When the nucleotide sequence of a variable region is compared between members of the same genotype, there is typically a great deal of similarity; however, there is typically very little similarity in the nucleotide sequence of the variable regions between members of different genotypes (see, for instance, Yamada et al., *Virology*, 223, 255-261 (1996)).

It is expected that a 5' NTR and a 3' NTR can be obtained from different sources, including molecularly cloned laboratory strains, for instance cDNA clones of HCV, and clinical isolates. Examples of molecularly cloned laboratory strains include the HCV that is encoded by pCV-H77C (Yanagi et al., *Proc. Natl. Acad. Sci. USA*, 94, 8738-8743 (1997), Genbank accession number AF011751, SEQ ID NO:1, where nucleotides 1-341 are the 5' NTR and nucleotides 9378-9599 are the 3' NTR), and pHCV-H (Inchauspe et al., *Proc. Natl. Acad. Sci. USA*, 88, 10292-10296 (1991), Genbank accession number M67463, SEQ ID NO:3, where nucleotides 1-341 are the 5' NTR and nucleotides 9378-9416 are the 3' NTR). Clinical isolates can be from a source of infectious HCV, including tissue samples, for instance from blood, plasma, serum, liver biopsy, or leukocytes, from an infected animal, including a human or a primate. It is also expected that the polynucleotide encoding the HCV polyprotein present in a replication competent polynucleotide can be prepared by recombinant, enzymatic, or chemical techniques.

In some aspects, a 5' NTR and a 3' NTR of a replication competent polynucleotide of the present invention is not a specific genotype. For instance, a 5' NTR and a 3' NTR present in a replication competent polynucleotide of the present invention can be genotype 1a, 1b, 1c, 2a, 2b, 2c, 3a, 3b, 4, 5a, or 6a (as defined by Simmons, *Hepatology*, 21, 570-583 (1995)). In other aspects, the HCV polyprotein is genotype 1a. Methods for determining the genotype of a 5' NTR and a 3' NTR are routine and known to the art and include evaluation of the

nucleotide sequence for specific nucleotides that are characteristic of a specific genotype.

In some aspects of the invention a replication competent polynucleotide includes a second coding region. The second coding sequence may be present in the 3'NTR, for instance, in the variable region of the 3'NTR. In some aspects of the invention, the second coding region is present in the variable region such that the variable region is not removed. Alternatively, the second coding region replaces the variable region in whole or in part. In some aspects of the invention, for instance, when the HCV has the genotype 1a, the second coding region is inserted in the variable region between nucleotides 5 and 6 of the sequence 5' CUCUUAAGC 3', where the sequence shown corresponds to the positive-strand.

In some aspects of the invention, the second coding region is present in a replication competent polynucleotide downstream of the 5'NTR, and upstream of the first coding region, i.e., the coding region encoding a HCV polyprotein. For instance, the first nucleotide of the second coding region may be immediately downstream and adjacent to the last nucleotide of the 5'NTR. Alternatively, the first nucleotide of the second coding region may be further downstream of the last nucleotide of the 5'NTR, for instance, about 2 to about 51 nucleotides, about 33 to about 51 nucleotides, or about 36 to about 48 nucleotides downstream of the last nucleotide of the 5'NTR. Typically, when the first nucleotide of the second coding region is not immediately downstream of the last nucleotide of the 5'NTR, the nucleotides in between the 5'NTR and the second coding region encode the amino terminal amino acids of the HCV core polypeptide. For instance, the 5' end of the second coding region may further include about 33 to about 51 nucleotides, or about 36 to about 48 nucleotides, that encode the first about 11 to about 17, or about 12 to about 16, amino acids of the core polypeptide. The result is a fusion polypeptide made up of amino terminal amino acids of the core polypeptide and the polypeptide encoded by the second coding region (see, for instance, Yi et al., Virol., 304, 197-210 (2002), and U.S. Published Application US 2003 0125541). Without intending to be limiting, it is believed the presence of the nucleotiodes from the core coding sequence act to enhance translation the polypeptide encoded by the second coding region.

In those aspects of the invention where the second coding region present in a replication competent polynucleotide is present downstream of the 5'NTR and

upstream of the coding region encoding the HCV polyprotein, the replication competent polynucleotide typically includes a regulatory region operably linked to the downstream coding region, e.g., the coding region encoding the HCV polyprotein. Preferably, the regulatory region provides for the translation of the downstream coding region. The size of the regulatory region may be from about 400 nucleotides to about 800 nucleotide, more preferably, about 600 nucleotides to about 700 nucleotides. Typically, the regulatory region is an IRES. Examples of IRES elements are described herein.

The second coding region can encode a polypeptide including, for instance, a marker, including a detectable marker and/or a selectable marker. Examples of detectable markers include molecules having a detectable enzymatic activity, for instance, secretory alkaline phosphatase, molecules having a detectable fluorescence, for instance, green or red or blue fluorescent protein, and molecules that can be detected by antibody. Examples of selectable markers include molecules that confer resistance to antibiotics able to inhibit the replication of eukaryotic cells, including the antibiotics kanamycin, ampicillin, chloramphinicol, tetracycline, blasticidin, neomycin, and formulations of phleomycin D1 including, for example, the formulation available under the trade-name ZEOCIN (Invitrogen, Carlsbad, California). Coding sequences encoding such markers are known to the art. Other examples of polypeptides that can be encoded by the second coding region include a transactivator, and/or a fusion polypeptide. Preferably, when the polypeptide is a fusion polypeptide, the second coding region includes nucleotides encoding a marker, more preferably, nucleotides encoding a fusion between a transactivator and a marker. Transactivators are described herein below. Optionally, the coding region can encode an immunogenic polypeptide. A replication competent polynucleotide containing a second coding region is typically dicistronic, i.e., the coding region encoding the HCV polyprotein and the second coding region are separate.

An "immunogenic polypeptide" refers to a polypeptide which elicits an immunological response in an animal. An immunological response to a polypeptide is the development in a subject of a cellular and/or antibody-mediated immune response to the polypeptide. Usually, an immunological response includes but is not limited to one or more of the following effects: the production

of antibodies, B cells, helper T cells, suppressor T cells, and/or cytotoxic T cells, directed specifically to an epitope or epitopes of the polypeptide fragment.

A transactivator is a polypeptide that affects in *trans* the expression of a coding region, preferably a coding region integrated in the genomic DNA of a cell. Such coding regions are referred to herein as "transactivated coding regions." The cells containing transactivated coding regions are described in detail herein below. Transactivators useful in the present invention include those that can interact with a regulatory region, preferably an operator sequence, that is operably linked to a transactivated coding region. As used herein, the term "transactivator" includes polypeptides that interact with an operator sequence and either prevent transcription from initiating at, activate transcription initiation from, or stabilize a transcript from, a transactivated coding region operably linked to the operator sequence. Examples of useful transactivators include the HIV tat polypeptide (see, for example, the polypeptides MEPVDPRLEPWKHPGSQPKTACTNCYCKKCCFHCQVCFITKALGISYGRK KRRQRRRAHQNSQTHQASLSKQPTSQPRGDPTGPKE (SEQ ID NO:5) which is encoded by nucleotides 5377 to 5591 and 7925 to 7970 of Genbank accession number AF033819), and MEPVDPRLEPWKHPGSQPKTACTNCYCKKCCFHCQVCFITKALGISYGRK KRRQRRRPPQGSQTHQVSLSKQPTSQSRGDPTGPKE (SEQ ID NO:10). The HIV tat polypeptide interacts with the HIV long terminal repeat (LTR). Other useful transactivators include human T cell leukemia virus tax polypeptide (which binds to the operator sequence tax response element, Fujisawa et al., *J. Virol.*, 65, 4525-4528 (1991)), and transactivating polypeptides encoded by spumaviruses in the region between env and the LTR, such as the bel-1 polypeptide in the case of human foamy virus (which binds to the U3 domain of these viruses, Rethwilm et al., *Proc. Natl. Acad. Sci. USA*, 88, 941-945 (1991)). Alternatively, a post-transcriptional transactivator, such as HIV rev, can be used. HIV rev binds to a 234 nucleotide RNA sequence in the env gene (the rev-response element, or RRE) of HIV (Hadzopolou-Cladaras et al., *J. Virol.*, 63, 1265-1274 (1989)).

Other transactivators that can be used are those having similarity with the amino acid sequence of SEQ ID NO:5 or SEQ ID NO:10. The similarity is generally determined as described herein above. A candidate amino acid sequence that is being compared to an amino acid sequence present in SEQ ID

NO:5 or SEQ ID NO:10 can be isolated from a virus, or can be produced using recombinant techniques, or chemically or enzymatically synthesized. Preferably, two amino acid sequences are compared using the Blastp program of the BLAST 2 search algorithm, as described herein above. Preferably, a transactivator includes an amino acid sequence having a structural similarity with SEQ ID NO:5 or SEQ ID NO:10, of at least about 90 %, at least about 94 %, at least about 96 %, at least about 97 %, at least about 98 %, or at least about 99 % identity. Typically, an amino acid sequence having a structural similarity with SEQ ID NO: 5 or SEQ ID NO:10 has tat activity. Whether such a polypeptide has activity can be evaluated by determining if the amino acid sequence can interact with an HIV LTR, preferably alter transcription from a coding sequence operably linked to an HIV LTR. Useful HIV LTRs are described herein.

Active analogs or active fragments of a transactivator can be used in the invention. An active analog or active fragment of a transactivator is one that is able to interact with an operator sequence and either prevent transcription from initiating at, activate transcription initiation from, or stabilize a transcript from, a transactivated coding region operably linked to the operator sequence.

Active analogs of a transactivator include polypeptides having conservative amino acid substitutions that do not eliminate the ability to interact with an operator and alter transcription. Substitutes for an amino acid may be selected from other members of the class to which the amino acid belongs. For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and tyrosine. Polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, aspartate, and glutamate. The positively charged (basic) amino acids include arginine, lysine, and histidine. The negatively charged (acidic) amino acids include aspartic acid and glutamic acid. Examples of preferred conservative substitutions include Lys for Arg and *vice versa* to maintain a positive charge; Glu for Asp and *vice versa* to maintain a negative charge; Ser for Thr so that a free -OH is maintained; and Gln for Asn to maintain a free NH_2.

Active fragments of a transactivator include a portion of the transactivator containing deletions or additions of about 1, about 2, about 3, about 4, or at least about 5 contiguous or noncontiguous amino acids such that the resulting transactivator will alter expression of an operably linked transactivated coding

region. A preferred example of an active fragment of the HIV tat polypeptide includes amino acids amino acids 1-48 of SEQ ID NO: 5, or amino acids 1-48 of SEQ ID NO:10.

In those aspects of the invention where the second coding region encodes a fusion polypeptide, the fusion polypeptide can further include amino acids corresponding to a *cis*-active proteinase. When the fusion polypeptide is a fusion between a transactivator and a marker, preferably the fusion polypeptide also includes amino acids corresponding to a *cis*-active proteinase. Preferably the amino acids corresponding to a *cis*-active proteinase are present between the amino acids corresponding to the transactivator and the marker. A *cis*-active proteinase in this position allows the amino acids corresponding to the transactivator and the marker to be physically separate from each other in the cell within which the replication competent polynucleotide is present. Examples of *cis*-active proteinases that are useful in the present invention include the *cis*-active 2A proteinase of foot-and-mouth disease (FMDV) virus (see, for example, US Patent 5,846,767 (Halpin et al.) and US Patent 5,912,167 (Palmenberg et al.)), ubiquitin (see, for example, Tauz et al., *Virology*, 197, 74-85 (1993)), and the NS3 recognition site GADTEDVVCCSMSY (SEQ ID NO:6) (see, for example, Lai et al., *J. Virol.*, 74, 6339-6347 (2000)).

Active analogs and active fragments of *cis*-active proteinases can also be used. Active analogs of a *cis*-acting proteinase include polypeptides having conservative amino acid substitutions that do not eliminate the ability of the proteinase to catalyze cleavage. Active fragments of a *cis*-active proteinase include a portion of the *cis*-active proteinase containing deletions or additions of one or more contiguous or noncontiguous amino acids such that the resulting *cis*-active proteinase will catalyze the cleavage of the proteinase.

In some aspects of the invention, the second coding region may further include an operably linked regulatory region. Preferably, a regulatory region located 5' of the operably linked coding region provides for the translation of the coding region.

A preferred regulatory region located 5' of an operably linked second coding region is an internal ribosome entry site (IRES). An IRES allows a ribosome access to mRNA without a requirement for cap recognition and subsequent scanning to the initiator AUG (Pelletier, et al., *Nature*, 334, 320-325

(1988)). An IRES is located upstream of the translation initiation codon, e.g., ATG or AUG, of the coding sequence to which the IRES is operably linked. The distance between the IRES and the initiation codon is dependent on the type of IRES used, and is known to the art. For instance, poliovirus IRES initiates a ribosome translocation/scanning process to a downstream AUG codon. For other IRES elements, the initiator codon is generally located at the 3' end of the IRES sequence. Examples of an IRES that can be used in the invention include a viral IRES, preferably a picornaviral IRES or a flaviviral IRES. Examples of poliovirus IRES elements include, for instance, poliovirus IRES, encephalomyocarditis virus IRES, or hepatitis A virus IRES. Examples of preferred flaviviral IRES elements include hepatitis C virus IRES, GB virus B IRES, or a pestivirus IRES, including but not limited to bovine viral diarrhea virus IRES or classical swine fever virus IRES. Other IRES elements with similar secondary and tertiary structure and translation initiation activity can either be generated by mutation of these viral sequences, by cloning of analogous sequences from other viruses (including picornaviruses), or prepared by enzymatic synthesis techniques.

The size of the second coding region is not critical to the invention. It is expected there is no lower limit on the size of the second coding region, and that there is an upper limit on the size of the second coding region. This upper limit can be easily determined by a person skilled in the art, as second coding region that are greater than this upper limit adversely affect replication of a replication competent polynucleotide. The second coding region is typically at least about 10 nucleotides, at least about 20 nucleotides, at least about 30 nucleotides, or at least about 40 nucleotides.

A replication competent polynucleotide may also include a nucleotide sequence having cis-acting ribozyme activity. Such a ribozyme is typically present at the 3' end of the 3' NTR of a replication competent polynucleotide, and generates a precise 3' terminal end of the replication competent polynucleotide when it is an RNA molecule by cleaving the junction between the replication competent polynucleotide and the ribozyme. This can be advantageous when the replication competent polynucleotide is to be used for a transient transfection. Since the ribozyme catalyzes its own removal from the RNA molecule, this type

of ribozyme is present only when a replication competent polynucleotide is a DNA molecule.

The replication competent polynucleotide of the invention can be present in a vector. When a replication competent polynucleotide is present in a vector the polynucleotide is DNA, including the 5' non-translated RNA and the 3' non-translated RNA, and, if present, the second coding sequence. Methods for cloning and/or inserting hepatitis C virus sequences into a vector are known to the art (see, e.g., Yanagi et al., *Proc. Natl. Acad. Sci., USA*, 94, 8738-8743 (1997); and Rice et al., (U.S. Patent 6,127,116)). Such constructs are often referred to as molecularly cloned laboratory strains, and an HCV that is inserted into a vector is often referred to as a cDNA clone of the HCV. If the RNA encoded by the HCV is able to replicate *in vivo*, the HCV present in the vector is referred to as an infectious cDNA clone. A vector is a replicating polynucleotide, such as a plasmid, phage, cosmid, or artificial chromosome to which another polynucleotide may be attached so as to bring about the replication of the attached polynucleotide. A vector can provide for further cloning (amplification of the polynucleotide), i.e., a cloning vector, or for expression of the polypeptide encoded by the coding region, i.e., an expression vector. The term vector includes, but is not limited to, plasmid vectors, viral vectors, cosmid vectors, or artificial chromosome vectors. Preferably the vector is a plasmid. Preferably the vector is able to replicate in a prokaryotic host cell, for instance *Escherichia coli.* Preferably, the vector can integrate in the genomic DNA of a eukaryotic cell.

An expression vector optionally includes regulatory sequences operably linked to the replication competent polynucleotide such that it is transcribed to produce RNA molecules. These RNA molecules can be used, for instance, for introducing a replication competent polynucleoitde into a cell that is in an animal or growing in culture. The terms "introduce" and "introducing" refer to providing a replication competent polynucleotide to a cell under conditions that the polynucleotide is taken up by the cell in such a way that it can then replicate. The replication competent polynucleotide can be present in a virus particle, or can be a nucleic acid molecule, for instance, RNA. The invention is not limited by the use of any particular promoter, and a wide variety are known. Promoters act as regulatory signals that bind RNA polymerase in a cell to initiate transcription of a downstream (3' direction) HCV. The promoter used in the invention can be a

constitutive or an inducible promoter. A preferred promoter for the production of replication competent polynucleotide as an RNA molecule is a T7 promoter.

The present invention includes methods for identifying a replication competent polynucleotide, including detecting and/or selecting for cells containing a replication competent polynucleotide. Typically, the cells used in this aspect of the invention are primate or human cells growing in culture. Useful cultured cells will support the replication of the polynucleotides of the present invention, and include primary human or chimpanzee hepatocytes, peripheral mononuclear cells, cultured human lymphoid cell lines (for instance lines expressing B-cell and T-cell markers such as Bjab and Molt-4 cells), and continuous cell lines derived from such cells, including HPBMa10-2 and Daudi (Shimizu et al., *J. Gen. Virol.*, 79, 1383-1386 (1998), and MT-2 (Kato et al., *Biochem. Biophys. Res. Commun.*, 206, 863-869 (1995)). Other useful cells include those derived from a human hepatoma cells, for instance, Huh-7 (see, for instance, Lohmann et al. (*Science*, 285, 110–113 (1999)), Huh-7.5 (see, for instance, Blight et a., *J. Virol.*, 76, 13001-13014 (2002), and Blight et al., *J. Virol.*, 77, 3183-3190 (2003)), HepG2 and IMY-N9 (Date et al., *J. Biol. Chem.*, 279, 22371-22376 (2004)), and PH5CH8 (Ikeda et al., *Virus Res.*, 56, 157-167 (1998)). In general, useful cells include those that support replication of HCV RNA, including, for instance, replication of the HCV encoded by pCV-H77C, replication of the HCV encoded by pHCV-N as modified by Beard et al. (*Hepatol.*, 30, 316-324 (1999)), or replication of such an HCV modified to contain one or more adaptive mutations.

In some aspects of the invention, the cultured cell includes a polynucleotide that includes a coding region, the expression of which is controlled by a transactivator. Such a coding region is referred to herein as a transactivated coding region. A transactivated coding region encodes a marker, such as a detectable marker, for example, secretory alkaline phosphatase (SEAP), an example of which is encoded by nucleotides 748-2239 of SEQ ID NO:7 (see Fig. 9). Typically, a cultured cell that includes a polynucleotide having a transactivated coding region is used in conjunction with a replication competent polynucleotide of the persent invention that includes a coding region encoding a transactivator.

The polynucleotide that includes the transactivated coding region can be present integrated into the genomic DNA of the cell, or present as part of a vector that is not integrated. Methods of modifying a cell to contain an integrated DNA are known to the art (see, for instance, Lemon et al., U.S. Published Application US 2003 0125541, and Yi et al., Virol., 302, 197-210 (2002)).

Operably linked to the transactivated coding region is an operator sequence. The interaction of a transactivator with an operator sequence can alter transcription of the operably linked transactivated coding region. In those aspects of the invention where a transactivator increases transcription, there is typically low transcription, or, essentially no transcription, of the transactivated coding region in the absence of a transactivator. An operator sequence can be present upstream (5') or downstream (3') of a transactivated coding region. An operator sequence can be a promoter, or can be a nucleotide sequence that is present in addition to a promoter.

In some aspects of the invention, the operator sequence that is operably linked to a transactivated coding sequence is an HIV long terminal repeat (LTR). An example of an HIV LTR is depicted at nucleotides 1-719 of SEQ ID NO:7. Also included in the present invention are operator sequences having similarity to nucleotides 1-719 of SEQ ID NO:7. The similarity between two nucleotides sequences may be determined by aligning the residues of the two polynucleotides (i.e., the nucleotide sequence of the candidate operator sequence and the nucleotide sequence of nucleotides 1-719 of SEQ ID NO:7) to optimize the number of identical nucleotides along the lengths of their sequences; gaps in either or both sequences are permitted in making the alignment in order to optimize the number of shared nucleotides, although the nucleotides in each sequence must nonetheless remain in their proper order. A candidate operator sequence can be isolated from a cell, or can be produced using recombinant techniques, or chemically or enzymatically synthesized. Preferably, two nucleotide sequences are compared using the Blastn program of the BLAST 2 search algorithm, as described by Tatusova, et al. (*FEMS Microbiol Lett* 1999, 174:247-250), and available at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. Preferably, the default values for all BLAST 2 search parameters are used, including reward for match = 1, penalty for mismatch = -2, open gap penalty = 5, extension gap penalty = 2, gap x_dropoff = 50, expect = 10, wordsize = 11, and filter on. In the comparison of

two nucleotide sequences using the BLAST search algorithm, structural similarity is referred to as "identities." Preferably, an operator sequence includes a nucleotide sequence having a structural similarity with the nucleotides 1-719 of SEQ ID NO:7 of at least about 90 %, at least about 95 %, or at least about 99 % identity. Typically, an operator sequence having structural similarity with the nucleotides 1-719 of SEQ ID NO:7 has transcriptional activity. Whether such an operator sequence has transcriptional activity can be determined by evaluating the ability of the operator sequence to alter transcription of an operably linked coding sequence in response to the presence of a polypeptide having tat activity, preferably, a polypeptide including the amino acids of SEQ ID NO:5 or SEQ ID NO:10.

A selecting agent may be used to inhibit the replication of cultured cells that support the replication of polynucleotides of the present invention. Examples of selecting agents include antibiotics, including kanamycin, ampicillin, chloramphenicol, tetracycline, neomycin, and formulations of phleomycin D1. A selecting agent can act to prevent replication of a cell, or kill a cell, while the agent is present and the cell does not express a molecule that provides resistance to the selecting agent. Typically, the molecule providing resistance to a selecting agent is expressed in the cell by a replication competent polynucleotide of the present invention. Alternatively, the molecule providing resistance to a selecting agent is expressed by the cell but the expression of the molecule is controlled by a replication competent polynucleotide of the present invention that is present in the cell. The concentration of the selecting agent is typically chosen such that a cell does not replicate if it does not contain a molecule providing resistance to a selecting agent. The appropriate concentration of a selecting agent varies depending on the particular selecting agent, and can be easily determined by one having ordinary skill in the art using known techniques.

When a polynucleotide is introduced into a cell that is growing in culture, the polynucleotide can be introduced using techniques known to the art. Such techniques include, for instance, liposome and non-liposome mediated transfection. Non-liposome mediated transfection methods include, for instance, electroporation.

In some aspects of the invention, when a replication competent polynucleotide is identified using cultured cells, its ability to replicate may be

verified by introducing the replication competent polynucleotide into a cell present in an animal, preferably a chimpanzee. When the cell is present in the body of an animal, the replication competent polynucleotide can be introduced by, for instance, subcutaneous, intramuscular, intraperitoneal, intravenous, or percutaneous intrahepatic administration, preferably by percutaneous intrahepatic administration. Methods for determining whether a replication competent polynucleotide is able to replicate in a chimpanzee are known to the art (see, for example, Yanagi et al., *Proc. Natl. Acad. Sci. USA*, 94, 8738-8743 (1997)). In general, the demonstration of infectivity is based on the appearance of the virus in the circulation of the chimpanzee over the days and weeks following the intrahepatic injection of the replication competent polynucleotide. The presence of the virus can be confirmed by reverse transcription-polymerase chain reaction (RT-PCR) detection of the viral RNA, by inoculation of a second chimpanzee with transfer of the hepatitis C virus infection as indicated by the appearance of liver disease and seroconversion to hepatitis C virus in ELISA tests, or possibly by the immunologic detection of components of the hepatitis C virus (e.g., the core protein) in the circulation of the inoculated animal. It should be noted that seroconversion by itself is generally not a useful indicator of infection in an animal injected with a viral RNA produced using a molecularly cloned laboratory strain, as this RNA may have immunizing properties and be capable of inducing HCV-specific antibodies to proteins translated from an input RNA that is non-replicating. Similarly, the absence of seroconversion does not exclude the possibility of viral replication and infection of a chimpanzee with HCV.

Whether a polynucleotide is replication competent can be determined using methods known to the art, including methods that use nucleic acid amplification to detect the result of increased levels of replication. For instance, transient transfection of a cell with a replication competent polynucleotide permits measurement of the production of additional polynucleotides. Methods for transient transfection of a cell with a replication competent polynucleotide and for assay of subsequent replication are known to the art. In some aspects of the invention, another method for detecting a replication competent polynucleotide includes measuring the production of viral particles by a cell. The measurement of viral particles can be accomplished by passage of supernatant from media containing a cell culture that may contain a replication competent polynucleotide,

and using the supernatant to infect a second cell. Detection of the polynucleotide or viral particles in the second cell indicates the initial cell contains a replication competent polynucleotide. The production of infectious virus particles by a cell can also be measured using antibody that specifically binds to an HCV viral particle. As used herein, an antibody that can "specifically bind" an HCV viral particle is an antibody that interacts only with the epitope of the antigen (e.g., the viral particle or a polypeptide that makes up the particle) that induced the synthesis of the antibody, or interacts with a structurally related epitope. "Epitope" refers to the site on an antigen to which specific B cells and/or T cells respond so that antibody is produced. An epitope could include about 3 amino acids in a spatial conformation which is unique to the epitope. Generally an epitope includes at least about 5 such amino acids, and more usually, consists of at least about 8-10 such amino acids. Antibodies to HCV viral particles can be produced as described herein.

In another aspect, identifying a replication competent polynucleotide includes incubating a cultured cell that includes a polynucleotide of the present invention. In those aspects of the invention where the replication competent polynucleotide includes a second coding region encoding a detectable marker, cells containing the replication competent polynucleotide can be identified by observing individual cells that contain the detectable marker. Alternatively, if the detectable marker is secreted by the cell, the presence of the marker in the medium in which the cell is incubated can be detected. Methods for observing the presence or absence of a detectable marker in a cell or in liquid media are known to the art.

Another aspect of the invention provides for the positive selection of cells that include a replication competent polynucleotide. In this aspect of the invention, a replication competent polynucleotide typically includes a second coding sequence encoding a selectable marker, and the cell which includes the replication competent polynucleotide is incubated in the presence of a selecting agent. Those cells that can replicate in the presence of the selecting agent contain a polynucleotide that is replication competent. The cells that can replicate are detected by allowing resistant cells to grow in the presence of the selecting agent, and observing, for instance, the presence of colonies and/or the expression of a marker, such as SEAP.

In some aspects, the method may further include isolating virus particles from the cells that contain a replication competent polynucleotide and exposing a second cell to the isolated virus particle under conditions such that the virus particle is introduced to the cell. After providing time for expression of the selectable marker, the second cell is then incubated with the selecting agent. The presence of a cell that replicates indicates the replication competent polynucleotide produces infectious virus particles.

In another aspect, the invention provides a method for detecting a replication competent polynucleotide. The method includes incubating a cell that contains a replication competent polynucleotide of the present invention. The polynucleotide may include a second coding region encoding a selectable or detectable marker. Optionally, the polynucleotide may include a transactivator that interacts with the operator sequence present in the cell . In this aspect, the cell may include a transactivated coding region and an operator sequence operably linked to the transactivated coding region. The method further includes detecting the presence of increased amounts of the replication competent polynucleotide, or the presence or absence of the marker encoded by the second coding sequence or the transactivated coding region present in the cell. The presence of increased amounts of the replication competent polynucleotide or the marker indicates the cell includes a replication competent polynucleotide.

The methods described above for identifying a replication competent polynucleotide can also be used for identifying a variant replication competent polynucleotide, i.e., a replication competent polynucleotide that is derived from a replication competent polynucleotide of the present invention. A variant replication competent polynucleotide may have a faster replication rate than the parent or input polynucleotide. The method takes advantage of the inherently high mutation rate of RNA replication. It is expected that during continued culture of a replication competent polynucleotide in cultured cells, the polynucleotide of the present invention may mutate, and some mutations will result in polynucleotides with greater replication rates. The method includes identifying a cell that has greater expression of a polypeptide encoded by a replication competent polynucleotide. A polynucleotide of the present invention that replicates at a faster rate will result in more of the polynucleotide in the cell, or will result in more of the polypeptide(s) that is encoded by the second coding

region present in the polynucleotide. For instance, when a replication competent polynucleotide encodes a selectable marker, a cell containing a variant polynucleotide having a greater replication rate will be resistant to higher levels of an appropriate selecting agent. When a polynucleotide encodes a transactivator, a cell containing a variant polynucleotide having a greater replication rate than the parent or input polynucleotide will express higher amounts of the transactivated coding region that is present in the cell.

A cDNA molecule of a variant replication competent polynucleotide can be cloned using methods known to the art (see, for instance, Yanagi et al., *Proc. Natl. Acad. Sci., USA*, 94, 8738-8743 (1997)). The nucleotide sequence of the cloned cDNA can be determined using methods known to the art, and compared with that of the input RNA. This allows identification of mutations that have occurred in association with passage of the replication competent polynucleotide in cell culture. For example, using methods known to the art, including longrange RT-PCR, extended portions of a variant replication competent polynucleotide genome can be obtained. Multiple clones could be obtained from each segment of the genome, and the dominant sequence present in the culture determined. Mutations that are identified by this approach can then be reintroduced into the background of the cDNA encoding the parent or input polynucleotide.

The present invention also provides methods for identifying a compound that inhibits replication of a replication competent polynucleotide. The method includes contacting a cell containing a replication competent polynucleotide with a compound and incubating the cell under conditions that permit replication of the replication competent polynucleotide in the absence of the compound. After a period of time sufficient to allow replication of the polynucleotide, the replication competent polynucleotide is detected. A decrease in the presence of replication competent polynucleotide in the cell contacted with the compound relative to the presence of replication competent polynucleotide in a cell not contacted by the compound indicates the compound inhibits replication of the polynucleotide. A compound that inhibits replication of such a polynucleotide includes compounds that completely prevent replication, as well as compounds that decrease replication. Preferably, a compound inhibits replication of a replication competent polynucleotide by at least about 50%, more preferably at least about 75%, most preferably at least about 95%.

The compounds added to a cell can be a wide range of molecules and is not a limiting aspect of the invention. Compounds include, for instance, a polyketide, a non-ribosomal peptide, a polypeptide, a polynucleotide (for instance an antisense oligonucleotide or ribozyme), other organic molecules, or a combination thereof. The sources for compounds to be screened can include, for example, chemical compound libraries, fermentation media of *Streptomycetes*, other bacteria and fungi, and extracts of eukaryotic or prokaryotic cells. When the compound is added to the cell is also not a limiting aspect of the invention. For instance, the compound can be added to a cell that contains a replication competent polynucleotide. Alternatively, the compound can be added to a cell before or at the same time that the replication competent polynucleotide is introduced to the cell.

Typically, the ability of a compound to inhibit replication of a replication competent polynucleotide is measured using methods described herein. For instance, methods that use nucleic acid amplification to detect the amount of a replication competent polynucleotidein a cell can be used. Alternatively, methods that detect or select for a marker encoded by a replication competent polynucleotide or encoded by a cell containing a replication competent polynucleotide can be used.

In some aspects of the invention, the replication competent polynucleotide of the invention can be used to produce viral particles. Preferably, the viral particles are infectious. For instance, a cell that includes a replication competent polynucleotide can be incubated under conditions that allow the polynucleotide to replicate, and the viral particles that are produced can be isolated using methods routine and known to the art. The viral particles can be used as a source of virus particles for various assays, including evaluating methods for inactivating particles, excluding particles from serum, identifying a neutralizing compound, and as an antigen for use in detecting anti-HCV antibodies in an animal. An example of using a viral particle as an antigen includes use as a positive-control in assays that test for the presence of anti-HCV antibodies.

For instance, the activity of compounds that neutralize or inactivate the particles can be evaluated by measuring the ability of the molecule to prevent the particles from infecting cells growing in culture or in cells in an animal. Inactivating compounds include detergents and solvents that solubilize the

envelope of a viral particle. Inactivating compounds are often used in the production of blood products and cell-free blood products. Examples of compounds that can be neutralizing include a polyketide, a non-ribosomal peptide, a polypeptide (for instance, an antibody), a polynucleotide (for instance, an antisense oligonucleotide or ribozyme), or other organic molecules. Preferably, a neutralizing compound is an antibody, including polyclonal and monoclonal antibodies, as well as variations thereof including, for instance, single chain antibodies and Fab fragments.

Viral particles produced by replication competent polynucleotide of the invention can be used to produce antibodies. Laboratory methods for producing polyclonal and monoclonal antibodies are known in the art (see, for instance, Harlow E. et al. *Antibodies: A laboratory manual* Cold Spring Harbor Laboratory Press, Cold Spring Harbor (1988) and Ausubel, R.M., ed. *Current Protocols in Molecular Biology* (1994)), and include, for instance, immunizing an animal with a virus particle. Antibodies produced using the viral particles of the invention can be used to detect the presence of viral particles in biological samples. For instance, the presence of viral particles in blood products and cell-free blood products can be determined using the antibodies.

The present invention further includes methods of treating an animal including administering neutralizing antibodies. The antibodies can be used to prevent infection (prophylactically) or to treat infection (therapeutically), and optionally can be used in conjunction with other molecules used to prevent or treat infection. The neutralizing antibodies can be mixed with pharmaceutically acceptable excipients or carriers. Suitable excipients include but are not limited to water, saline, dextrose, glycerol, ethanol, or the like and combinations thereof. In addition, if desired, neutralizing antibodies and pharmaceutically acceptable excipients or carriers may contain minor amounts of auxiliary substances such as wetting or emulsifying agents, pH buffering agents, and/or adjuvants which enhance the effectiveness of the neutralizing antibodies. Such additional formulations and modes of administration as are known in the art may also be used.

The virus particles produced by replication competent polynucleotide of the invention can be used as a source of viral antigen to measure the presence and amount of antibody present in an animal. Assays are available that measure the

presence in an animal of antibody directed to HCV, and include, for instance, ELISA assays and recombinant immunoblot assay. These types of assays can be used to detect whether an animal has been exposed to HCV, and/or whether the animal may have an active HCV infection. However, these assays do not use virus particles, but rather individual or multiple viral polypeptides expressed from recombinant cDNA that are not in the form of virus particles. Hence they are generally unable to detect potentially important antibodies directed against surface epitopes of the envelope polypeptides, nor are they typically measures of functionally important viral neutralizing antibodies. Such antibodies are generally detected with the use of infectious virus particles, such as those that are produced in this system. The use of infectious viral particles as antigen in assays that detect the presence of specific antibodies by virtue of their ability to block the infection of cells with HCV viral particles, or that possibly bind to whole virus particles in an ELISA assay or radioimmunoassay, will allow the detection of functionally important viral neutralizing antibodies.

The present invention also provides a kit for identifying a compound that inhibits replication of a replication competent polynucleotide. The kit includes a replication competent polynucleotide as described herein, and a cell that contains a polynucleotide including a transactivated coding sequence encoding a detectable marker and an operator sequence operably linked to the transactivated coding sequence in a suitable packaging material. Optionally, other reagents such as buffers and solutions needed to practice the invention are also included. Instructions for use of the packaged materials are also typically included.

As used herein, the phrase "packaging material" refers to one or more physical structures used to house the contents of the kit. The packaging material is constructed by well known methods, preferably to provide a sterile, contaminant-free environment. The packaging material may include a label which indicates that the replication competent polynucleoitde can be used for identifying a compound that inhibits replication of such a polynucleotide. In addition, the packaging material may contain instructions indicating how the materials within the kit are employed. As used herein, the term "package" refers to a solid matrix or material such as glass, plastic, and the like, capable of holding within fixed limits the replication competent virus and the vertebrate cell.

The present invention is illustrated by the following examples. It is to be understood that the particular examples, materials, amounts, and procedures are to be interpreted broadly in accordance with the scope and spirit of the invention as set forth herein.

EXAMPLES

Materials and Methods

Cells. Huh7 cells were grown in Dulbecco's modified Eagle's medium (Gibco BRL, Carlsbad, CA) supplemented with 10% fetal calf serum, penicillin and streptomycin. En5-3 is a clonal cell line derived from Huh7 cells by stable transformation with the plasmid pLTR-SEAP (Yi et al., Virology, 304,197-210 (2002)). These cells were cultured in Dulbecco's modified Eagle's medium (Gibco BRL) supplemented with 10% fetal calf serum, 2 µg/ml blasticidin (Invitrogen), penicillin and streptomycin. Cell lines were passaged once or twice per week. G418 at a concentration of 250 µg/ml was used to select colonies from En5-3 cells transfected with replicon RNAs containing 1a sequences.

Plasmids. The plasmid pBpp-Htat2ANeo was constructed by replacing the *BsrG*I-*Xba*I fragment of pBpp-Ntat2ANeo/SI (identical to Ntat2ANeo/SI as described by Yi et al. (Yi et al., Virology, 304,197-210 (2002)) with the analogous segment of pH77c (GenBank AF011751) (Yanagi et al., Proc Natl Acad Sci USA, 94, 8738-43 (1997)) engineered to contain a *BsrG*I site at the corresponding location by Quick-Change (Stratagene, La Jolla, CA) mutagenesis. This fragment swap results in the NS3-NS5B sequence in pBpp-Htat2ANeo being identical to that of pH77c, with the exception of the RNA encoding the N-terminal 75 amino acid residues of NS3 that retains the genotype 1b Con1 sequence. Since Bpp-Ntat2ANeo/SI was originally engineered to contain the genotype 1a 5' nontranslated RNA (5'NTR) sequence (Yi et al., Virology, 304,197-210 (2002)), the resulting pBpp-Htat2ANeo construct possesses both a genotype 1a 5'NTR and 1a 3'NTR sequence. Overlapping PCR was used to fuse an anti-genomic hepatitis delta ribozyme sequence directly to the 3' end of the genotype 1a 3'NTR, in order to generate a self-cleaving 3' sequence with the exact 3' terminal nucleotide of HCV (Perrotta and Been, Nucleic Acids Res, 24,1314-21 (1996)). Derivatives of

pBpp-Htat2ANeo containing the adaptive mutations K1691R or S2204I were created by Quick-Change (Stratagene) mutagenesis.

To construct pBpp-H34A-Ntat2ANeo/SI, an *EcoR*I restriction site was created in pBpp-Ntat2ANeo/SI near the 3' end of the NS4A coding region by Quick-Change mutagenesis. After digestion of the resulting plasmid with *BsrG*I and *EcoR*I, the excised HCV segment was replaced with the equivalent sequence from pH77c which had been amplified by PCR using primers pairs containing terminal *BsrG*I and *EcoR*I sites, respectively. To construct the plasmid Hpp-H34A-Ntat2ANeo, DNA fragments representing the encephalomyocarditis virus (EMCV) internal ribosome entry site (IRES) sequence and the genotype 1a H77c NS3 protein-coding sequence were fused by overlapping PCR. The resulting fragment was digested with *Kpn*I at a site located within the EMCV IRES and *BsrG*I at the site created within the modified pH77c NS3 region (see above), then inserted in place of the corresponding fragment in pBpp-H34A-Ntat2ANeo/SI. The adaptive mutations, Q1067R or G1188R, were introduced into pHpp-H34A-Ntat2ANeo/SI in a similar fashion, using cDNA fragments prepared by RT-PCR of template RNAs isolated from independent G418-resistant replicon cell lines selected after transfection of En5-3 cells with Hpp-H34A-Ntat2ANeo RNA. pHtat2ANeo/SI was constructed by replacing the *BsrG*I-*Xba*I fragment of pHpp-H34A-Ntat2ANeo/SI with that of pBpp-Htat2ANeo/SI. A similar strategy was used to construct pHtat2ANeo/QR/SI, pHtat2ANeo/KR/SI, and pHtat2ANeo/QR/KR/SI. Quick-Change (Stratagene) mutagenesis was used to introduce the P1496L, F2080V and K2040R mutations into replicon constructs derived from pHtat2ANeo/SI.

Modified pH77c plasmids containing adaptive mutations were created by replacing the *BsrG*I-*Xba*I fragment with the corresponding fragment from the pHtat2ANeo plasmid derivative containing the indicated mutation, except for the Q1067R mutation which was introduced by Quick-Change (Stratagene) mutagenesis. Each mutation was confirmed by sequence analysis. For use as controls, replication-incompetent subgenomic and genome-length genotype 1a constructs (Htat2ANeo/QR/VI/KR/KR5A/SI/AAG and H77/QR/VI/KR/KR5A/SI/AAG) were created by replacing residues 2737-2739 of NS5B ('GDD') with 'AAG' using a similar strategy. Each mutation was confirmed by sequence analysis.

RNA transcription and transfection. RNA was synthesized with T7 MEGAScript reagents (Ambion, Austin, TX), after linearizing plasmids with *Xba*I. Following treatment with RNase-free DNase to remove template DNA and precipitation of the RNA with lithium chloride, the RNA was transfected into Huh7 cells or En5-3 cells by electroporation. Briefly, 5 μg RNA was mixed with 2 x 10^6 cells suspended in 500 μl phosphate buffered saline, in a cuvette with a gap width of 0.2 cm (Bio-Rad). Electroporation was with two pulses of current delivered by the Gene Pulser II electroporation device (Bio-Rad), set at 1.5 kV, 25 μF, and maximum resistance. For transient replication assays, no G418 was added to the media. Transfected cells were transferred to two wells of a 6-well tissue culture plate, and culture medium removed completely every 24 hrs and saved at 4°C for subsequent SEAP assay. The cells were washed twice with PBS prior to re-feeding with fresh culture medium. Since the culture medium was replaced every 24 hours in these transient assays, the SEAP activity measured in these fluids reflected the daily production of SEAP by the cells. Cells were split 5 days after transfection. Samples of media were stored at 4°C until assayed for SEAP activity at the conclusion of the experiment.

Alkaline phosphatase assay. SEAP activity was measured in 10 μl aliquots of transfected cell supernatant culture fluids using the Phospha-Light Chemiluminescent Reporter Assay (Applied Biosystems/Tropix, Foster City, CA) with the manufacturer's suggested protocol reduced in scale. The luminescent signal was read using a TD-20/20 Luminometer (Turner Designs, Inc., Sunnyvale, CA).

Sequence analysis of cDNA from replicating HCV RNAs. HCV RNA was extracted from cells, converted to cDNA and amplified by PCR as described previously (Yi et al., J Virol,77, 57-68 (2003)). First-strand cDNA synthesis was carried out with Superscript II reverse transcriptase (Gibco-BRL); pfu-Turbo DNA polymerase (Stratagene) was used for PCR amplification of the DNA. The amplified DNAs were subjected to direct sequencing using an ABI 9600 automatic DNA sequencer.

In vitro translation. In vitro transcribed RNA, prepared as described above, was used to program in vitro translation reactions in rabbit reticulocyte lysate (Promega, Madison, WI). Approximately 1μg RNA, 2 μl of [^{35}S]-methionine (1,000 Ci/mmol at 10 mCi/ml), and 1 μl of an amino acid mixture

lacking methionine were included in each 50 μl reaction mixture. Translation was carried out at 30° C for 90 minutes. Translation products were separated by SDS-PAGE followed by autoradiography or PhosphorImager (Molecular Dynamics) analysis.

Indirect immunofluorescence. Cells were grown on chamber slides until 70-80% confluent, washed 3 times with PBS, and fixed in methanol/acetone (1:1 V/V) for 10 min at room temperature. A 1:20 dilution of a primary, murine monoclonal antibody to core or NS5A (Maine Biotechnology Services, Portland, ME) was prepared in PBS containing 3% bovine serum albumin, and incubated with the fixed cells for 1 hour at room temperature. Following additional washes with PBS, specific antibody binding was detected with a goat anti-mouse IgG FITC-conjugated secondary antibody (Sigma, St. Louis, Missouri) diluted 1:70. Cells were washed with PBS, counterstained with DAPI, and mounted in Vectashield mounting medium (Vector Laboratories, Burlingame, CA) prior to examination by a Zeiss AxioPlan2 Fluorescence microscope.

Northern analysis for HCV RNA. Replicon-bearing cells were seeded into 10 cm dishes at a density of $5x10^5$ cells/dish, and harvested the RNA 4 days later. Total cellular RNA was extracted with Trizol reagent (Gibco-BRL) and quantified by spectrophotometry at 260 nm. Thirty μg of the total RNA extracted from each well was loaded onto a denaturing agarose-formaldehyde gel, subjected to electrophoresis and transferred to positively-charged Hybond-N+ nylon membranes (Amersham-Pharmacia Biotec) using reagents provided with the NorthernMax Kit (Ambion). RNAs were immobilized on the membranes by UV-crosslinking. The membrane was hybridized with a mixture of [^{32}P]-labeled antisense riboprobe complementary to the 3'-end of the HCV NS5B sequence (nucleotides 8990-9275) derived from pH77C or pHCV-N, and the hybridized probe was detected by exposure to X-ray film.

Results

Transient replication of 1a replicon containing chimeric NS3-coding sequence. In contrast to genotype 1b HCV, several previous reports suggest that it is difficult to generate subgenomic genotype 1a replicons that are capable of efficient replication in Huh7 cells (Blight et al., Science, 290:1972-4 (2000), Guo et al., J Virol, 75, 8516-23 (2001), Ikeda et al., J Virol, 76, 2997-3006 (2002),

Lanford et al., J Virol, 77,1092-104 (2003)). Similar results were encountered with a dicistronic SEAP reporter replicon constructed from the H77c infectious molecular clone (Yanagi et al., Proc Natl Acad Sci USA, 94, 8738-43 (1997)) that encoded both the HIV tat protein and neomycin phosphotransferase in the upstream cistron. The organization of this latter replicon, Htat2ANeo/SI (Fig. 1), was similar to that of the efficiently replicating, genotype 1b Bpp-Ntat2ANeo/SI replicon (Fig. 1), referred to previously simply as "Ntat2ANeo/SI" (Yi et al., Virology, 304,197-210 (2002)). Most of the HCV polyprotein-coding sequence in Bpp-Ntat2ANeo/SI was derived from the genotype 1b HCV-N strain of HCV (Beard et al., Hepatol., 30, 316-24 (1999)), but the "Bpp" prefix used here and throughout this communication refers to the presence of 225 nucleotides (nts) of sequence that are derived from the Con1 strain of HCV at the extreme 5' end of the polyprotein coding region ("pp" indicates the 5' proximal protease-coding region, Fig. 1). In contrast, all of the HCV sequence in Htat2ANeo/SI (Fig. 1) is derived from the genotype 1a H77c virus, including both the 5' NTR and 3' NTR sequences. Unlike Bpp-Ntat2ANeo/SI RNA, Htat2ANeo/SI RNA did not transduce the selection of G418-resistant colonies, nor induce secretion of SEAP above that observed with a replication-incompetent NS5B-deletion mutant (ΔGDD) when transfected into En5-3 cells (stably transformed Huh7 cells that express SEAP under control of the HIV long terminal repeat promoter) (Yi et al., Virology, 304,197-210 (2002)). in a transient replication assay. This was the case even though the replicon was engineered to contain the genotype 1b adaptive mutation, S2204I, within NS5A (Fig. 1). The absence of apparent replication of Htat2ANeo/SI RNA was striking given the fact that it was derived from a well-documented infectious molecular clone of the H77c strain of HCV (Yanagi et al., Proc Natl Acad Sci USA, 94, 8738-43 (1997)).

Recent reports suggest that the EMCV IRES-driven translation of the second cistron in dicistronic, subgenomic RNAs such as those shown in Fig. 1 may be reduced when the translated RNA sequence is derived from genotype 1a virus, rather than genotype 1b (Gu et al., J Virol, 77, 5352-9 (2003), Guo et al., J Virol, 75, 8516-23 (2001), Lanford et al., J Virol, 77,1092-104 (2003)). However, even when translation of the second cistron is rendered more efficient by replacing the 5' 225 nts of the genotype 1a NS3 sequence with related sequence from the Con1 genotype 1b virus, replication typically has not been observed when the

remainder of the replicon sequence is derived from a genotype 1a virus (Guo et al., J Virol, 75, 8516-23 (2001), Lanford et al., J Virol, 77,1092-104 (2003)). However, Gu et al. (Gu et al., J Virol, 77, 5352-9 (2003)) recently described the successful selection of a replication competent, chimeric replicon in which the 5' 225 nts of the NS3 coding sequence was derived from genotype 1b virus, and the remainder of the second cistron from genotype 1a HCV (construction of chimeric replicons being simplified by a unique *BsrG*1 site within the genotype 1b Con1 virus sequence, 225 nts downstream from the 5' end of the NS3 region). This replicon also contained 5'NTR sequence derived from genotype 1b virus, and had a single base change within the genotype 1a 3'NTR sequence. The results of Gu et al. (Gu et al., J Virol, 77, 5352-9 (2003)) suggest that the inclusion of the Con1 sequence at the 5' end of the NS3 region may in some way facilitate replication of the 1a RNA. This hypothesis is strengthened by observations made with genotype 1b replicons derived from HCV-N. Those described previously, including Bpp-Ntat2ANeo/SI RNA, were constructed by ligation of HCV-N sequence to a Con1 replicon at the *BsrG*1 site (Guo et al., J Virol, 75, 8516-23 (2001), Ikeda et al., J Virol, 76, 2997-3006 (2002), Yi et al., Virology, 304,197-210 (2002)), and thus they contain 5' proximal NS3 sequence (proximal protease sequence or 'pp', Fig. 1) derived from the Con1 virus. Although this chimeric Con1/HCV-N RNA replicates significantly more efficiently than the originally-described Con1 replicons, the replacement of the 5' proximal NS3 sequence in Bpp-Ntat2ANeo/SI with sequence from HCV-N (resulting in Npp-Ntat2ANeo/SI) virtually ablated its replication phenotype in transient transfection assays, although it remained possible to select G418-resistant colonies at a low frequency following transfection.

To formally assess the ability of the 5' proximal genotype 1b NS3 sequence to enhance genotype 1a RNA replication, the 5' 225 nts of NS3 coding region in Htat2ANeo/SI were replaced with the Con1 sequence, generating Bpp-Htat2ANeo/SI (Fig. 1). The construct was also modified by replacing the *Xba*1 restriction site at the 3' end of the HCV sequence with the hepatitis delta virus ribozyme sequence (Perrotta and Been, Nucleic Acids Res, 24,1314-21 (1996)). We have shown previously that the presence of the 4 extraneous nts at the 3' end of the replicon RNA that results from run-off transcription of *Xba*l-digested plasmid DNA reduces the replication competence of genotype 1b RNAs by 2-3

fold (Yi and Lemon, Rna, 9, 331-45 (2003)). The inclusion of the ribozyme resulted in self-cleaving RNA transcripts capable of generating the exact 3' terminal HCV RNA sequence. Nonetheless, this modified Bpp-Htat2ANeo/SI RNA still remained incapable of inducing the expression of SEAP in transfected EN5-3 cells beyond that observed following transfection of the ΔGDD RNA. Transfection resulted only in an initial burst in SEAP expression due to translation of the input replicon RNA, without the sustained SEAP expression that is indicative of RNA replication (Fig. 2). However, the Bpp-Htat2ANeo/SI RNA was capable of transducing the selection of G418-resistant cell colonies supporting replication of the RNA over a period of 3-4 weeks following transfection of the cells.

The sequence of replicon RNAs extracted from two independent G418-resistant cell clones selected following the transfection of En5-3 cells with Bpp-Htat2ANeo RNA was analyzed. The presence of a single Lys to Arg mutation located within the NS4A region, at residue 1691 (K1691R) of the polyprotein in both cell clones was determined. This residue is located just beyond the 3' limits of the NS4A cofactor peptide sequence which participates in forming a noncovalent complex with NS3 and enhances its protease activity (Wright-Minogue et al., J Hepatol, 32, 497-504 (2000), Yao et al., Structure Fold Des, 7, 1353-63 (1999)). To determine whether the K1691R mutation facilitated replication of the chimeric genotype 1b/1a RNA in En5-3 cells, this mutation was introduced into the parental Bpp-Htat2ANeo/SI construct, thereby creating Bpp-Htat2ANeo/KR/SI (see Table 2 for a list of all adaptive mutations identified in these studies, as well as the symbols used to indicate their presence in constructs). As shown in Fig. 2, this single mutation significantly enhanced the replication capacity of the RNA, allowing replication to be detected by a sustained increase in SEAP expression following transient transfection of EN5-3 cells in the absence of G418 (Fig. 2). Since the level of SEAP production has been shown to correlate closely with intracellular replicon RNA abundance in this reporter system (Yi et al., Virology, 304,197-210 (2002), Yi et al., J Virol,77, 57-68 (2003)) we conclude that K1691R is an adaptive mutation. Interestingly, this mutation has been shown previously to confer an enhanced replication phenotype on Con1 replicons (Lohmann et al., J Virol, 77, 3007-19 (2003)), whereas the sequence of HCV-N is naturally Arg at this position (Beard et al., Hepatol., 30, 316-24

(1999)). Our results stand in contrast to those reported by Gu et al. (Gu et al., J Virol, 77, 5352-9 (2003)), who identified several mutations within the NS3, NS5A, and NS5B sequences of chimeric genotype 1b-1a RNAs. None of these mutations appeared to enhance the ability of the chimeric RNA to replicate or transduce colony selection.

The 5' 225 nts of the genotype 1a NS3 sequence down modulate replicon amplification. The results described above, as well as those of Gu et al. (Gu et al., J Virol, 77, 5352-9 (2003)) suggest that first 225 nts of the genotype 1a NS3 sequence have a negative impact on the replication of subgenomic HCV replicons. This could occur by down modulation of EMCV IRES-directed translation of the nonstructural proteins (Guo et al., J Virol, 75, 8516-23 (2001)), or by directly influencing replication itself, possibly by influencing an NS3-related function. To address this issue, the identification of additional adaptive mutations capable of compensating for the presence of the 5' proximal genotype 1a protease sequence was sought. Thus additional chimeric replicons containing the entire genotype 1a NS3/4A sequence within the background of Bpp-Ntat2ANeo/SI (Hpp-H34A-Ntat2ANeo/SI, Fig. 3A) were constructed. Also constructed was a variant of this construct in which the first 225 nts of the NS3/4A sequence was replaced with Con1 sequence (Bpp-H34A-Ntat2ANeo/SI, Fig. 3A). In both chimeric RNAs, the sequence extending from NS4B to the 3'NTR was derived entirely from the genotype 1b HCV-N strain. While the replicon containing the entire genotype 1a NS3/4A sequence (Hpp-H34A-NtatNeo/SI) did not show evidence of replication in a transient transfection assay, the variant containing the first 225 nts of the Con1 sequence (Bpp-H34A-NtatNeo/SI) replicated as well as the reference Bpp-Ntat2ANeo/SI replicon (Fig. 3B). This result confirms that the 5' 225 nucleotides of the genotype 1a NS3 sequence have a negative effect on RNA replication in En5-3 cells, and also indicates that the downstream genotype 1a NS3/4A sequence functions well in this context.

Interestingly, despite the lack of detectable RNA replication in the transient assay, selection of stable G418-resistant cell clones following transfection of Hpp-H34A-Ntat2ANeo/SI RNA was possible. Sequencing of replicon RNAs derived from two independent cell clones revealed only a single potentially adaptive mutation in each: Q1067R and G1188R, both of which are located within RNA encoding the NS3 protease (Fig. 3A). The Q1067R mutation

is of particularly interest, since it is within the 5' 225 nucleotides of the NS3 region. When introduced into Hpp-H34A-Ntat2ANeo/SI, both the Q1067R and (to a lesser extent) the G1188R mutations enhanced replication of the RNA to a level that was detectable in the transient assay (Fig. 3B), indicating that both are adaptive mutations and capable of compensating, in part, for the presence of the genotype 1a protease sequence. However, neither of these mutations, when introduced into a replicon containing only genotype 1a sequence (Htat2ANeo/SI), was able to enhance replication to the point where it was evident in the transient assay (Htat2ANeo/QR/SI, Fig. 4).

Transient replication of a genotype 1a replicon in normal Huh7 cells. To determine whether the K1691R and Q1067R mutations might work cooperatively to confer a transient replication phenotype on the genotype 1a replicon RNA, both were introduced into Htat2ANeo and assessed the ability of the modified RNA to replicate in transfected En5-3 cells. Surprisingly, the combination of the K1691R and Q1067R mutations (in addition to the S2204I mutation in NS5A) conferred a relatively robust replication phenotype on the genotype 1a RNA, such that replication was easily detectable in the transient transfection assay using the SEAP reporter system (Htat2ANeo/QR/KR/SI, Fig. 4B). Using an approach similar to that taken in the preceding experiments, an additional adaptive mutation (F2080V) within the NS5A-coding region (F2080V) was subsequently identified, when cells transfected with Htat2ANeo/QR/KR/SI RNA were subjected to G418 selection pressure. This mutation resulted in slightly greater replication efficiency when introduced into the genotype 1a replicon containing K1691R and Q1067R in addition to S2204I (Htat2ANeo/QR/KR/FV/SI, Fig. 4B). However, F2080V had relatively little effect when added to replicons containing only K1691R or Q1067R (in addition to S2204I) (Fig. 4B). Minimally increased secretion of SEAP above the ΔGDD background was observed during the first 5 days after transfection with Htat2ANeo/KR/FV/SI, but this was no longer apparent after 6 days. The replication phenotype of Htat2ANeo/QR/FV/SI was indistinguishable from that of the replication incompetent ΔGDD mutant in this assay (Fig. 4B). These results are summarized in Fig. 4C.

To facilitate a comparison of these results with those reported previously by Blight et al. (Blight et al., J Virol, 77, 3181-90 (2003)), the adaptive P1496L mutation identified by this group within the helicase domain of NS3 following

transfection of a genotype 1a replicon was introduced into the highly permissive Huh7 subline, Huh-7.5. Consistent with the previous report, a 1a replicon bearing this mutation P1496L demonstrated only minimal evidence of replication in the transient assay (which utilizes En5-3 cells that are comparable to normal Huh7 cells in terms of their permissiveness for HCV RNA replication) (Htat2ANeo/PL/SI, Fig. 4B). The addition of the NS5A mutation, F2080V, failed to noticeably enhance the replication capacity of this RNA (Htat2ANeo/PL/FV/SI, Fig. 4B). SEAP expression induced by genotype 1a replicons containing both Q1067R and K1691R was approximately 10-fold that induced by replicons containing P1496L. Since SEAP production from En5-3 cells correlates closely with the intracellular abundance of replicon RNA (Yi et al., Virology, 304,197-210 (2002)), these results suggest that the protease domain mutations make a greater contribution to replication competence of the genotype 1a replicon.

Adaptive mutations within NS3 do not affect EMCV IRES-driven translation of the second cistron. As mentioned above, previous reports indicate that the EMCV-driven translation of the second cistron is reduced in genotype 1a replicons in comparison to replicons containing the genotype 1b Con1 sequence (Gu et al., J Virol, 77, 5352-9 (2003), Guo et al., J Virol, 75, 8516-23 (2001), Lanford et al., J Virol, 77,1092-104 (2003)). Although the mechanism is uncertain, the effect appears to be due to the genotype 1a sequence encoding the amino terminus of NS3. Since the adaptive Q1067R mutation is located within this region, we asked whether it or other mutations that enhance 1a replicon amplification do so by improving EMCV IRES-driven translation of the HCV nonstructural proteins. To test this hypothesis, in vitro translation reactions were programmed with genotype 1b and 1a replicon RNAs containing various adaptive mutations, and compared the production of proteins encoded by the second cistron with neomycin phosphotransferase produced from the first cistron. As shown in Fig. 5, the synthesis of NS3 was modestly reduced with replicons containing genotype 1a H77c sequence in the 5' proximal protease region (compare NS3 abundance in lanes 4-8 with that in other lanes). However, it was not increased by any of the adaptive mutations, including Q1067R. This result indicates that the difficulty of establishing replication competent 1a replicons is more likely due to the intrinsic property of the 1a sequence, than to an incompatibility of the HCV and EMCV sequences in this region leading to reduced activity of the EMCV

IRES. Nonetheless, the reduced level of translation of the genotype 1a nonstructural proteins that is evident in Fig. 5 may contribute to the poor replication phenotype of these RNAs.

An additional adaptive NS5A mutation further augments replication competence. Although the F2080V mutation in NS5A provided only a slight additional replication advantage to subgenomic genotype 1a RNAs containing the Q1067R, K1691R and S2204I mutations (Fig. 4), additional mutations were subsequently identified concurrently near the C-terminus of NS3 (V1655I) and within NS5A (K2040R) in RNAs replicating within a G418-resistant cell line selected following transfection with the subgenomic Htat2ANeo/QR/KR/SI replicon. As shown in Fig. 6, both of these mutations enhanced the replication capacity of genotype 1a RNA. Addition of the V1655I mutation resulted in a modest enhancement of Htat2ANeo/QR/KR/SI replication, leading to a replication phenotype slightly better that observed with the addition of the F2080V mutation. In contrast, the addition of the K2040R mutation in NS5A resulted in a dramatic increase in replication competence, rendering the replication phenotype of the genotype 1a RNA equivalent to that of the standard genotype 1b HCV-N replicon used in these studies, Bpp-Ntat2ANeo/SI (Fig. 6B). A genotype 1a replicon containing both of these adaptive mutations in addition to those identified earlier replicated with slightly greater efficiency than this reference genotype 1b RNA in the transient assay (Fig. 6B, Htat2ANeo/QR/VI/KR/KR5A/SI). These results were confirmed in independent experiments.

Robust replication of genome-length genotype 1a RNA with adaptive mutations. Encouraged by the above results, we assessed the in vitro replication competence of genomelength, genotype 1a H77c RNA engineered to contain the adaptive mutations described above. As with the dicistronic, subgenomic RNAs, we placed the hepatitis delta ribozyme sequence at the 3' end of the cloned infectious cDNA sequence in pH77c in order to generate RNA transcripts containing an exact HCV 3' terminus. As these genomic RNAs encoded no selectable marker or reporter protein product, their replication was assessed in transfected Huh7 and En5-3 cells by northern blot analysis in comparison with related subgenomic RNAs. Subgenomic and genome-length replication-incompetent H77 mutant RNAs, in which the GDD motif had been replaced with

AAG, served as negative controls for this experiment. For En5-3 cells transfected with the subgenomic RNAs, we also determined levels of SEAP expression.

As expected, the unmodified H77c RNA showed no evidence of replication, even though it has been shown previously to be infectious in chimpanzees when inoculated into liver (Fig. 7, compare lane 7 with the replication defective 1a genomic RNA in lane 11). The introduction of the Q1067R (NS3) mutation, alone or in combination with S2204I (NS5A), was insufficient to confer a detectable level of replication in Huh7 cells. However, when all three mutations were introduced (Q1067R, K1691R and S2204I), the H77c RNA acquired a relatively efficient replication phenotype with readily detectable amplification of the RNA in northern blots of cell lysates prepared 4 days after transfection of either Huh7 or En5-3 cells (Fig. 7, lane 8). Replication of the genome-length RNA was slightly increased by the further addition of the F2080V (NS5A) mutation (Fig. 7, lane 9). However, consistent with the data presented in Fig. 6, the inclusion of both the V1655I mutation in NS3 and the K2040R mutation conferred a substantially more robust replication phenotype on genome-length H77c, when present in combination with other adaptive mutations in NS3, NS4A and NS5A (H77c/QR/VI/KR/KR5A/SI, Fig. 7, compares lane10 and 11). This experiment thus confirmed the adaptive effects of these mutations. Northern blotting indicated that the replication capacity of genome-length genotype 1a RNAs containing adaptive mutations was significantly greater than the comparable subgenomic, dicistronic genotype 1a replicons, for which the RNA signal 4 days after transfection was low and near the limits of detection in northern blots (Fig. 7, compare lanes 3 to 6 with lanes 8 to 11). These findings are consistent with those reported previously by Blight et al. (*J. Virol.*, 77, 3181-3190 (2003)), and indicate that the inclusion of heterologous sequences in the dicistronic replicons impairs RNA replication competence. Subgenomic replicon RNA was detected unambiguously only in cells transfected with Htat2ANeo/QR/VI/KR/KR5A/SI, the RNA that generated the highest level of SEAP expression (Fig. 7, compare lane 5 and 6).

As a further measure of the replication competence of these modified genome-length H77c RNAs, we also examined transfected En5-3 cells for the presence of core or NS5A proteins using an indirect immunofluorescence method. Introduction of both the K1691R (NS4A) and S2204I mutations resulted in

detectable antigen expression 4 days after transfection, albeit only in a very low percentage of cells (less than 0.01%). However, strong expression of both the core and NS5A proteins was observed in approximately 30% of En5-3 cells 4 days after transfection of RNA containing all four adaptive mutations. Increased replication efficiency of genotype 1a RNAs correlated with a greater proportion of cells supporting the replication of HCV RNA, evidenced by the presence of viral antigen.

Discussion

Subgenomic, dicistronic, selectable HCV RNA replicons derived from genotype 1b viruses replicate efficiently in cultured cells (Blight et al, Science, 290:1972-1974 (2000), Guo et al., J. Virol., 75:8516-8523 (2001), Ikeda et al., J. Virol., 76:2997-3006 (2002), Krieger et al., J. Virol., 75:4614-4624 (2001), Lohmann et al., J. Virol., 75:1437-1449 (2001), and Lohmann et al., Science 285:110-113 (1999)). These novel RNAs have facilitated the study of HCV RNA replication and substantially accelerated antiviral drug discovery efforts. The Huh7 cell line, derived from a human hepatoma, appears to be uiquely permissive and supportive of the replication of these HCV RNAs, although recent studies suggest that other types of cells may also be permissive for HCV RNA replication (Zhu et al., J. Virol., 77:9204-9210 (2003)). However, despite the success of genotype 1b replicons, it has been difficult to generate RNAs that replicate efficiently in any cell type from other genotypes of HCV, including genotype 1a, (Blight et al, Science, 290:1972-1974 (2000), Guo et al., J. Virol., 75:8516-8523 (2001), Ikeda et al., J. Virol., 76:2997-3006 (2002), and Lanford et al., J. Virol., 77:1092-104 (2003)). This surprising observation indicates that significant biological differences exist between genotype 1a and 1b viruses, despite the fact that the nucleotide sequences of genotype 1a viruses are relatively closely related to those of genotype 1b (~90-93% identity). This biological difference raises the likelihood that antiviral agents that are found to be active against the genotype 1b virus may have significantly lesser activity against genotype 1a viruses. Considering these observations and the relatively high genetic variability that exists between different HCV genotypes, the development of cell culture systems supporting replication of viral RNAs from other genotypes will be important for validating in vitro efficacy of candidate antiviral agents across a range of

genetically distinct HCV genotypes, as well as developing a better overall understanding of these viruses.

Genotype 1a viruses are the most prevalent types of HCV in the United States, and like genotype 1 b virus they are relatively refractory to treatment with interferon (Fried et al., N Engl J Med, 347, 975-82 (2002), McHutchison and Fried, Clin Liver Dis, 7, 149-61 (2003)). Thus far, a detectable level of genotype 1a RNA replication has been reported only in specially isolated, highly permissive Huh7 human hepatoma cell sublines (e.g., Huh-7.5 cells) generated by eliminating the replication of genotype 1b RNA replicons from established replicon cell lines using interferon-α in vitro (Blight et al., J Virol, 77, 3181-90 (2003), Grobler et al., J Biol Chem, 278,16741-6 (2003)). These previously described genotype 1a RNAs possess cell culture-adaptive mutations that enhance their replication in these special cells, including those selected during the isolation of antibiotic-resistant cell lines containing these 1a replicons (Blight et al., J Virol, 77, 3181-90 (2003), Grobler et al., J Biol Chem, 278,16741-6 (2003)). However, the published reports suggest that these previously described genotype 1a RNAs do not replicate to a detectable level in standard Huh7 cells, and that their capacity for replication in cultured cells is thus limited. In contrast, genotype 1a HCV RNAs are reported here that replicate in a highly efficient manner in normal Huh7 cells.

Our results suggest that the highly efficient replication of genotype 1a RNAs requires at least three adaptive mutations located within the NS3, NS4A and NS5A proteins. It is evident that these mutations are mutually reinforcing in their ability to enhance the replication of the genotype 1a RNAs, even though they were identified individually under different circumstances. It was found that the introduction of the S2204I mutation in NS5A, which is known to promote the replication of genotype 1b virus RNAs in Huh7 cells (Blight et al., Science, 290:1972-4 (2000)), was not sufficient for subgenomic replicons composed entirely of the genotype 1a sequence to initiate replication in Huh7 cells. However, it made possible the selection of G418-resistant cell colonies following transfection of a chimeric replicon RNA, in which sequence from the infectious molecular clone of the genotype 1a H77c virus encoded all of the nonstructural proteins other than the N-terminal 75 amino acid residues of NS3 which were derived from the genotype 1b Con1 sequence (Fig. 1, Bpp-Htat2ANeo/SI). The HCV RNAs replicating in these cells contained a single mutation within the

NS4A-coding region (K1691R) that enhanced the replication capacity of the original chimeric replicon RNA (Fig. 2). These results suggest that a restriction to the replication of genotype 1a virus in Huh7 cells may reside within the serine protease domain of NS3, since substitution of the N-terminal domain of the genotype 1a protease with that from the Con1 genotype 1b virus allowed the initiation of replication and the selection of G418-resistant cells. A similar conclusion can be drawn from the results reported by Gu et. al. (Gu et al., J Virol, 77, 5352-9 (2003)). Thus, it is interesting that the adaptive mutation K1691R resides within NS4A very close to the surface of the NS3/4A protease complex that it helps to form (Fig. 8).

In an effort to better understand this restriction, a second chimeric replicon containing the complete genotype 1a NS34A sequence within the background of a genotype 1b replicon was constructed. This RNA (Hpp-H34A-Ntat2ANeo/SI) did not undergo detectable replication in the transient transfection system utilized in these studies (Fig. 3). However, it was capable of transducing the selection of G418-resistant cell colonies following transfection and antibiotic selection. Analysis of the sequence of the HCV RNAs replicating within these cells identified a second, cell culture-adaptive mutation within the N-terminal region of the NS3 protease (Q1067R), providing further evidence that a primary restriction to replication of genotype 1a virus resides within this domain. Yet additional evidence for this comes from the replication phenotype of the Bpp-H34A-Ntat2ANeo/SI replicon, which also contains all of the genotype 1a NS3/4A sequence except for the N-terminal 75 amino acid resides, and which demonstrated a robust replication phenotype in the transient transfection assay. Thus there appears to be no restriction to replication deriving from inclusion of the genotype 1a NS3 helicase domain, nor for that matter any part of the protease domain except for its N-terminus.

Further work demonstrated that the K1691R and Q1067R mutations worked cooperatively: neither by itself was capable of conferring the capacity for efficient replication on a replicon composed entirely of genotype 1a sequence, but a combination of the two (in addition to the genotype 1b S2204I adaptive mutation) resulted in RNA replication that could be readily detected in the transient transfection assay (Fig. 4). That these mutations should act cooperatively in their effects on replication, as indicated by the data shown in Fig. 4, is

consistent with their location in the polyprotein, since the NS4A protease cofactor domain interacts primarily with residues within the N-terminal domain of the NS3 protease (Wright-Minogue et al., J Hepatol, 32, 497-504 (2000), Yao et al., Structure Fold Des, 7, 1353-63 (1999)).

Additional adaptive mutations were identified and verified through an iterative series of experiments involving RNA transfection, isolation of G418-resistant cells, and analysis of the sequence of efficiently replicating genotype 1a RNAs. Also demonstrated was that the S2204I mutation did indeed facilitate the replication of the genotype 1a RNA, as its removal from the efficiently replicating subgenomic RNAs substantially reduced their replication competence in the transient transfection assay. The genotype 1a adaptive mutations identified herein are summarized in Table 2. They can be grouped functionally into two groups: K2040R, F2080V, and S2204I, which are all located within NS5A (a common site of genotype 1b adaptive mutations), and Q1067R, G1188R, V1655I, and K1691R, which are all located in or otherwise associated with the protease domain of NS3. While to some extent solvent exposed, both G1188R and Q1067R are close to the active site of the protease (Fig. 8), and would both add a significant charge difference to the active face of the protein. V1655I is particularly interesting. It is located near the extreme C-terminus of the NS3 protein, downstream of the helicase domain, and close to the protease active site in the crystal structure of the NS3/4A complex (Yao et al., Structure Fold Des, 7, 1353-63 (1999)). In the P3 position of the NS3/4A cleavage site, V1655 is certain to play a role in substrate recognition during the *cis*-active cleavage of the polyprotein at the NS3/4A junction and it remains within the substrate-binding pocket in the crystal structure. The potential impact of the K1691R mutation, within NS4A, on the conformation of the protease active site is much less certain, but it is in close proximity to the NS4A cofactor domain, as mentioned above, and intercalation of this domain into the NS3 protease is well known to modulate the activity of the protease.

Significantly, all of these NS3 and NS4A mutations are located at some distance from other genotype 1a adaptive mutations in NS3 that have been described in the literature (see Fig. 8). These mutations, located at S1222, A1226 and P1496, are all within the helicase domain of NS3 (Blight et al., J Virol, 77, 3181-90 (2003), Grobler et al., J Biol Chem, 278,16741-6 (2003)). While on the surface of the protein, they are located on the side opposite the solvent exposed

surfaces containing the G1188, V1655, and Q1067 residues (Fig. 8). Thus, it is possible that they facilitate genotype 1a RNA replication by a different mechanism than those mutations that cluster near the active site of the protease. At least the P1496L mutation identified by both Blight et al. (Blight et al., J Virol, 77, 3181-90 (2003)) and Grobler et al. (Grobler et al., J Biol Chem, 278,16741-6 (2003)) appears to be substantially less active in conferring replication capacity on the genotype 1a H77c RNA. This was demonstrated by the lack of detectable replication of RNA replicons containing this mutation (Htat2ANeo/PL/SI and Htat2ANeo/PL/FV/SI) in the transient transfection experiment summarized in Fig. 4.

What role could mutations near the active site of the NS3 protease play in promoting the replication of genotype 1a HCV RNA in Huh7 cells? It is unlikely that these mutations work by enhancing translation of the nonstructural proteins under control of the EMCV IRES in the context of the subgenomic replicon, since we observed no difference in translation of these proteins in vitro in reticulocyte lysates programmed with these RNAs (Fig. 5). More importantly, they enhance the replication of genomic H77c RNA lacking any heterologous sequence in Huh7 cells (see Fig. 7). These mutations do not seem likely to promote replication by favorably influencing the ability of the protease to process the viral polyprotein, since the polyprotein segment expressed in the Htat2ANeo derivatives is derived entirely from the same H77c genome, and this replicates very efficiently in chimpanzee liver. However, this does remain a formal possibility that needs to be excluded in future studies. It is possible, instead, that these mutations promote interactions of the NS3/4A complex with specific cellular proteins that play a role in assembly of the viral replicase complex, or otherwise influence replication by disabling innate cellular antiviral defenses.

Foy et al. (Foy et al., Science, 300, 1145-8 (2003)) recently demonstrated that expression of the NS3/4A protease effectively blocked activation of interferon regulatory factor 3 (IRF3) in Huh7 cells infected with Sendai virus, thereby preventing the induction of synthesis of interferon-β and other antiviral cytokines. This immuno-evasive action of NS3 was reversed by a specific ketoamide inhibitor of the NS3/4A protease, and was dependent upon the protease activity of NS3/4A, indicating that NS3/4A is likely to cleave a cellular protein involved in IRF3 signaling following viral infection. While Foy et al. (Foy et al., Science,

300, 1145-8 (2003)) demonstrated that both genotype 1a and genotype 1b proteases are capable of blocking IRF3 activation, it is intriguing to consider that the adaptive mutations within NS3/4A may promote its ability to direct such a cleavage, thereby enhancing replication of the virus by lessening cellular antiviral defenses.

The second group of adaptive mutations identified within NS5A, K2040R, F2080V, and S2204I (Table 2), are likely to function in a fashion similar to NS5A adaptive mutations identified in genotype 1b replicons, which include S2204I. Although their specific mechanism of action is not known, they may either promote the ability of NS5A to assemble a functional replicase complex in Huh7 cells, or perhaps augment the immunomodulatory actions that have been proposed for this viral protein through its interactions with double-stranded RNA stimulated protein kinase R (PKR) (Gale et al., Clin Diagn Virol, 10,157-62 (1998)). The contribution of these adaptive mutations to the replication of the genotype 1a RNA in these studies appears to be additive to that of the NS3/4A mutations (Figs. 3 and 6), not synergistic as shown for the combination of Q1067R and K1691R (Fig. 3).

The complete disclosure of all patents, patent applications, and publications, and electronically available material (including, for instance, nucleotide sequence submissions in, e.g., GenBank and RefSeq, and amino acid sequence submissions in, e.g., SwissProt, PIR, PRF, PDB, and translations from annotated coding regions in GenBank and RefSeq) cited herein are incorporated by reference. The foregoing detailed description and examples have been given for clarity of understanding only. No unnecessary limitations are to be understood therefrom. The invention is not limited to the exact details shown and described, for variations obvious to one skilled in the art will be included within the invention defined by the claims.

All headings are for the convenience of the reader and should not be used to limit the meaning of the text that follows the heading, unless so specified.

Sequence Listing Free Text

SEQ ID NO:1 Nucleotide sequence of Hepatitis C virus strain H77

SEQ ID NO:2 Amino acid sequence of HCV polyprotein encoded by nucleotides 342 - 9377 of SEQ ID NO:1.

SEQ ID NO:3 Nucleotide sequence of Hepatitis C virus strain H

SEQ ID NO:4 Amino acid sequence of HCV polyprotein encoded by nucleotides 342 - 9377 of SEQ ID NO:3.

SEQ ID NO:5 HIV tat polypeptide

SEQ ID NO:6 NS3 recognition site

SEQ ID NO:7 Nucleotide sequence of HIV SEAP, HIV long terminal repeat (LTR) is depicted at nucleotides 1-719, and secretory alkaline phosphatase is encoded by the nucleotides 748-2239.

SEQ ID NO:8 Nucleotide sequence of a 3' NTR.

SEQ ID NO:9 Nucleotide sequence of a 5' NTR

SEQ ID NO:10 HIV tat polypeptide

SEQ ID NO:11 genomic length hepatitis C virus, genotype 1a

SEQ ID NO:12 HCV polyprotein encoded by the coding region present in SEQ ID NO:11.

SEQ ID NO:13 nucleotide sequence of Htat2ANeo

SEQ ID NO:14 HCV polyprotein encoded by the coding region present in SEQ ID NO:13.

What is claimed is:

1. A replication competent polynucleotide comprising:

a 5' non-translated region (NTR), a 3' NTR, and a first coding sequence present between the 5' NTR and 3' NTR and encoding a hepatitis C virus polyprotein, wherein the polyprotein comprises an isoleucine at about amino acid 2204, and further comprises an adaptive mutation selected from the group of an arginine at about amino acid 1067, an arginine at about amino acid 1691, valine at about amino acid 2080, an isoleucine at about amino acid 1655, an arginine at about amino acid 2040, an arginine at about amino acid 1188, and a combination thereof.

2. The replication competent polynucleotide of claim 1 further comprising a second coding sequence.

3. The replication competent polynucleotide of claim 2 wherein the second coding sequence encodes a marker.

4. The replication competent polynucleotide of claim 2 wherein the second coding sequence encodes a transactivator.

5. The replication competent polynucleotide of claim 1 wherein the 5' NTR, the 3' NTR, and the nucleotide sequence encoding the polyprotein are genotype 1a.

6. The replication competent polynucleotide of claim 1 wherein the hepatitis C virus polyprotein is a subgenomic hepatitis C virus polyprotein.

7. The replication competent polynucleotide of claim 1 wherein the hepatitis C virus polyprotein comprises cleavage products core, E1, E2, P7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B.

8. The replication competent polynucleotide of claim 1 further comprising a nucleotide sequence having cis-acting ribozyme activity, wherein the nucleotide sequence is located 3' of the 3' NTR.

9. A replication competent polynucleotide comprising:

a 5' non-translated region (NTR), a 3' NTR, and a first coding sequence present between the 5' NTR and 3' NTR and encoding a hepatitis C virus polyprotein comprising cleavage products core, E1, E2, P7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B, wherein the polyprotein comprises an isoleucine at about amino acid 2204, an arginine at about amino acid 1067, and an arginine at about amino acid 1691.

10. The polynucleotide of claim 9 further comprising a second coding sequence.

11. The polynucleotide of claim 10 wherein the second coding sequence encodes a marker .

12. The polynucleotide of claim 10 wherein the second coding sequence encodes a transactivator.

13. The polynucleotide of claim 9 wherein the 5' NTR, polyprotein, and 3' NTR are genotype 1a.

14. A method for making a replication competent polynucleotide comprising:

providing a polynucleotide comprising a 5' NTR, 3' NTR, a first coding sequence present between the 5' NTR and 3' NTR and encoding a hepatitis C virus polyprotein, wherein the polyprotein comprises a serine at about amino acid 2204, a glutamine at about amino acid 1067, a lysine at about amino acid 1691, a phenylalanine at about amino acid 2080, a valine at about amino acid 1655, a lysine at about amino acid 2040, or a glycine at about amino acid 1188 and wherein the 5' NTR, polyprotein, and 3' NTR are genotype 1a; and

altering the coding sequence such that the polyprotein encoded thereby comprises an isoleucine at amino acid 2204, and at least one adaptive mutation

selected from the group consisting of an arginine at about amino acid 1067, an arginine at about amino acid 1691, a valine at about amino acid 2080, an isoleucine at about amino acid 1655, an arginine at about amino acid 2040, an arginine at about amino acid 1188, and a combination thereof.

15. The method of claim 14 wherein the hepatitis C virus polyprotein is a subgenomic hepatitis C virus polyprotein.

16. The method of claim 14 wherein the hepatitis C virus polyprotein comprises cleavage products core, E1, E2, P7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B.

17. A replication competent polynucleotide produced by the method of claim 14.

18. A method for identifying a compound that inhibits replication of a replication competent polynucleotide, the method comprising:

contacting a cell comprising a replication competent polynucleotide with a compound, the replication competent polynucleotide comprising 5' NTR, 3' NTR, a first coding sequence present between the 5' NTR and 3' NTR and encoding a hepatitis C virus polyprotein, wherein the hepatitis C virus polyprotein comprises an isoleucine at about amino acid 2204, and further comprises an adaptive mutation selected from the group of an arginine at about amino acid 1067, an arginine at about amino acid 1691, valine at about amino acid 2080, an isoleucine at about amino acid 1655, an arginine at about amino acid 2040, an arginine at about amino acid 1188, and a combination thereof;

incubating the cell under conditions wherein the replication competent polynucleotide replicates in the absence of the compound; and

detecting the replication competent polynucleotide, wherein a decrease of the replication competent HCV polynucleotide in the cell contacted with the compound compared to the replication competent polynucleotide in a cell not contacted with the compound indicates the compound inhibits replication of the replication competent polynucleotide.

19. The method of claim 18 wherein detecting the replication competent polynucleotide comprises nucleic acid amplification.

20. The method of claim 18 wherein the replication competent polynucleotide further comprises a second coding sequence encoding a marker, and wherein detecting the replication competent polynucleotide comprises identifying the marker.

21. The method of claim 18 wherein the replication competent polynucleotide further comprises a second coding sequence encoding a transactivator, wherein the cell comprises a polynucleotide comprising a transactivated coding sequence encoding a detectable marker and an operator sequence operably linked to the transactivated coding sequence, wherein the transactivator interacts with the operator sequence and alters expression of the transactivated coding sequence, and wherein detecting the replication competent polynucleotide in the cell comprises detecting the detectable marker encoded by the transactivated coding sequence.

22. The method of claim 18 wherein the cell is a human hepatoma cell.

23. The method of claim 18 wherein the hepatitis C virus polyprotein is a subgenomic hepatitis C virus polyprotein.

24. The method of claim 18 wherein the hepatitis C virus polyprotein comprises cleavage products core, E1, E2, P7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B.

25. The method of claim 18 wherein the 5' NTR, polyprotein, and 3' NTR are genotype 1a.

26. A method for selecting a replication competent polynucleotide, the method comprising:

incubating a cell in the presence of a selecting agent, wherein:

the cell comprises a polynucleotide comprising a 5' non-translated region (NTR), a 3' NTR, and a first coding sequence present between

the 5'NTR and 3'NTR and encoding a hepatitis C virus polyprotein, and a second coding sequence, wherein the polyprotein comprises an isoleucine at about amino acid 2204, and further comprises an adaptive mutation selected from the group of an arginine at about amino acid 1067, an arginine at about amino acid 1691, valine at about amino acid 2080, an isoleucine at about amino acid 1655, an arginine at about amino acid 2040, an arginine at about amino acid 1188, and a combination thereof;

the second coding sequence encodes a selectable marker conferring resistance to the selecting agent; and

the selecting agent inhibits replication of a cell that does not express the selectable marker; and

detecting a cell that replicates in the presence of the selecting agent, wherein the presence of such a cell indicates the polynucleotide is replication competent.

27. The method of claim 26 wherein the selecting agent is an antibiotic.

28. The method of claim 26 wherein the cell is a human hepatoma cell.

29. The method of claim 26 wherein the cell is a first cell, the method further comprising:

obtaining a virus particle produced by the first cell;

exposing a second cell to the isolated virus particle and incubating the second cell in the presence of the selecting agent; and

detecting a second cell that replicates in the presence of the selecting agent, wherein the presence of such a cell indicates the replication competent polynucleotide in the first cell produces an infectious virus particle.

30. The method of claim 26 wherein the hepatitis C virus polyprotein is a subgenomic hepatitis C virus polyprotein.

31. The method of claim 26 wherein the hepatitis C virus polyprotein comprises cleavage products core, E1, E2, P7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B.

32. The method of claim 26 wherein the 5'NTR, polyprotein, and 3'NTR are genotype 1a.

33. A method for detecting a replication competent polynucleotide, the method comprising:

incubating a cell comprising a replication competent polynucleotide, wherein:

the replication competent polynucleotide comprises a 5' non-translated region (NTR), a 3'NTR, and a first coding sequence present between the 5'NTR and 3'NTR and encoding a hepatitis C virus polyprotein, and a second coding sequence encoding a transactivator, wherein the polyprotein comprises an isoleucine at about amino acid 2204, and further comprises an adaptive mutation selected from the group of an arginine at about amino acid 1067, an arginine at about amino acid 1691, valine at about amino acid 2080, an isoleucine at about amino acid 1655, an arginine at about amino acid 2040, an arginine at about amino acid 1188, and a combination thereof;

the cell comprises a transactivated coding region and an operator sequence operably linked to the transactivated coding region; and

the transactivated coding region encodes a detectable marker, wherein the transactivator alters transcription of the transactivated coding region; and

detecting the detectable marker, wherein the presence of the detectable marker indicates the cell comprises a replication competent polynucleotide.

34. The method of claim 33 wherein the hepatitis C virus polyprotein is a subgenomic hepatitis C virus polyprotein.

35. The method of claim 33 wherein the hepatitis C virus polyprotein comprises cleavage products core, E1, E2, P7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B.

36. The method of claim 33 wherein the 5' NTR, polyprotein, and 3' NTR are genotype 1a.

37. A kit comprising:

a replication competent polynucleotide comprising a 5' non-translated region (NTR), a 3' NTR, and a first coding sequence present between the 5' NTR and 3' NTR and encoding a hepatitis C virus polyprotein, and a second coding sequence encoding a transactivator, wherein the polyprotein comprises an isoleucine at about amino acid 2204, and further comprises an adaptive mutation selected from the group of an arginine at about amino acid 1067, an arginine at about amino acid 1691, valine at about amino acid 2080, an isoleucine at about amino acid 1655, an arginine at about amino acid 2040, an arginine at about amino acid 1188, and a combination thereof; and

a cell comprising a polynucleotide comprising a transactivated coding sequence encoding a detectable marker and an operator sequence operably linked to the transactivated coding sequence, wherein the transactivator interacts with the operator sequence and alters expression of the transactivated coding sequence.

38. The method of claim 37 wherein the hepatitis C virus polyprotein is a subgenomic hepatitis C virus polyprotein.

39. The method of claim 37 wherein the hepatitis C virus polyprotein comprises cleavage products core, E1, E2, P7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B.

40. The method of claim 37 wherein the 5' NTR, polyprotein, and 3' NTR are genotype 1a.

Fig. 1

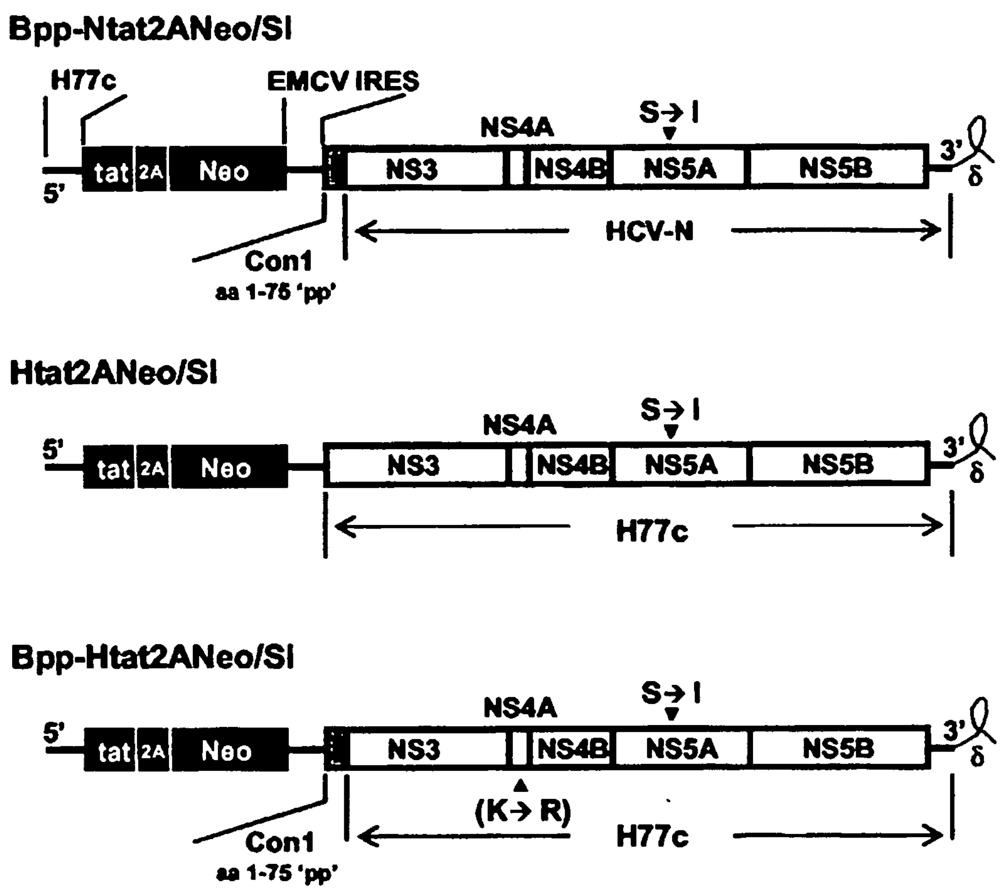

Bpp-Ntat2ANeo/SI
H77c
EMCV IRES
NS4A
S→I
5'
tat
2A
Neo
NS3
NS4B
NS5A
NS5B
3'
δ
HCV-N
Con1
aa 1-75 'pp'
Htat2ANeo/SI
NS4A
S→I
5'
tat
2A
Neo
NS3
NS4B
NS5A
NS5B
3'
δ
H77c
Bpp-Htat2ANeo/SI
NS4A
S→I
5'
tat
2A
Neo
NS3
NS4B
NS5A
NS5B
3'
δ
(K→R)
H77c
Con1
aa 1-75 'pp'

Fig. 2

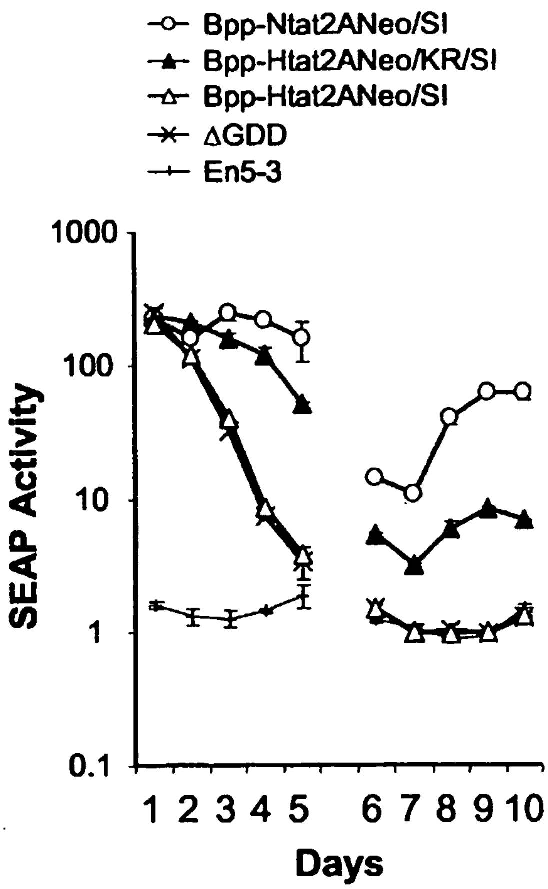

Bpp-Ntat2ANeo/SI
Bpp-Htat2ANeo/KR/SI
Bpp-Htat2ANeo/SI
ΔGDD
En5-3
1000
100
10
1
0.1
SEAP Activity
1 2 3 4 5 6 7 8 9 10
Days

Fig. 3A

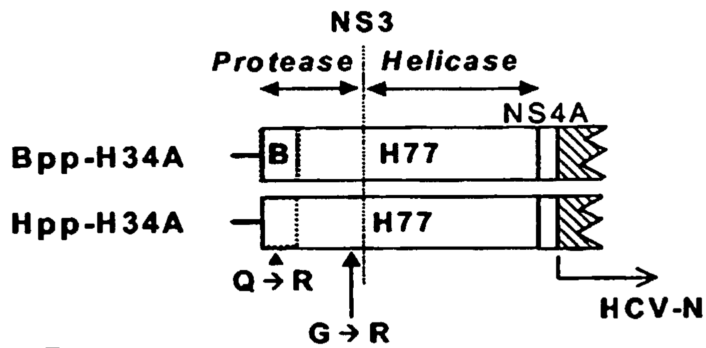

NS3
Protease
Helicase
NS4A
Bpp-H34A
B
H77
Hpp-H34A
H77
Q → R
G → R
HCV-N

Fig. 3B

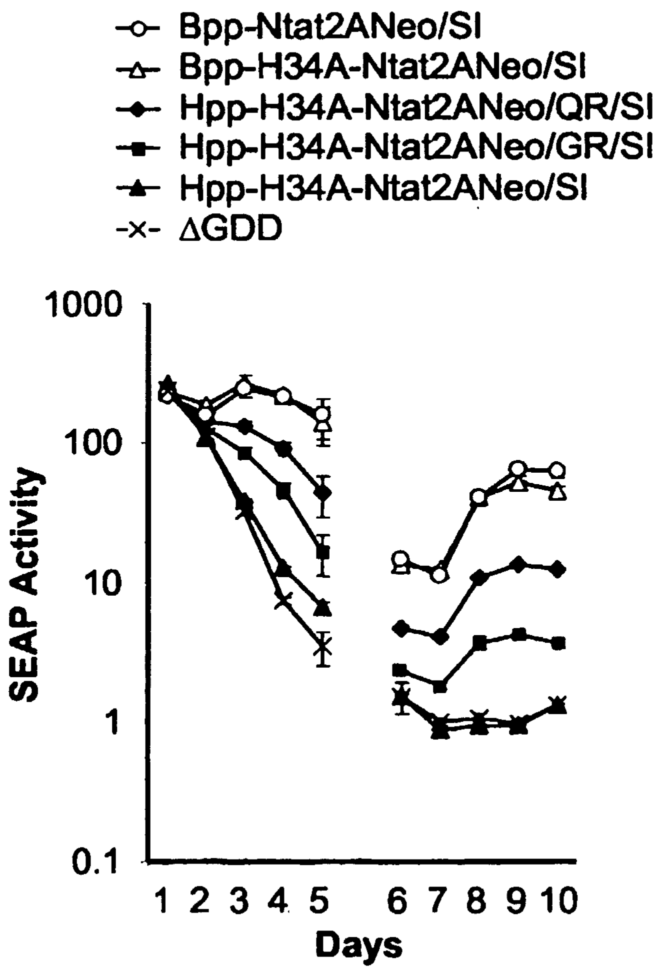

Bpp-Ntat2ANeo/SI
Bpp-H34A-Ntat2ANeo/SI
Hpp-H34A-Ntat2ANeo/QR/SI
Hpp-H34A-Ntat2ANeo/GR/SI
Hpp-H34A-Ntat2ANeo/SI
ΔGDD
SEAP Activity
1000
100
10
1
0.1
1 2 3 4 5 6 7 8 9 10
Days

Fig. 4A

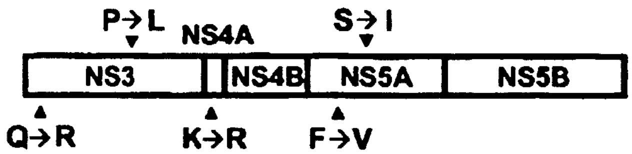


Fig. 4B

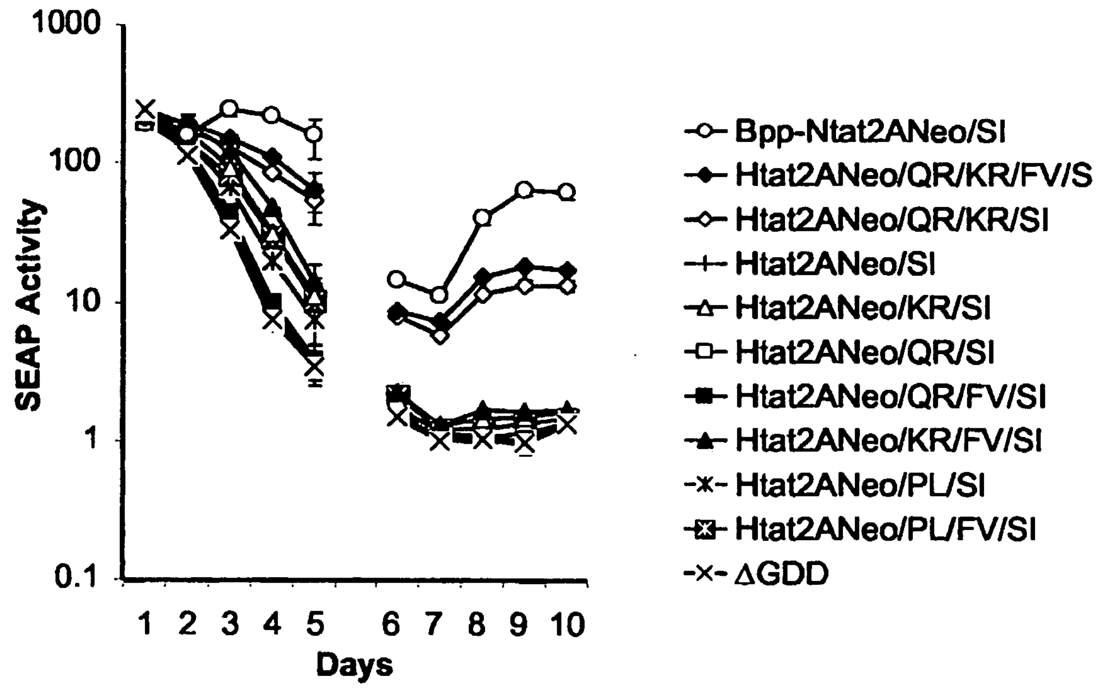
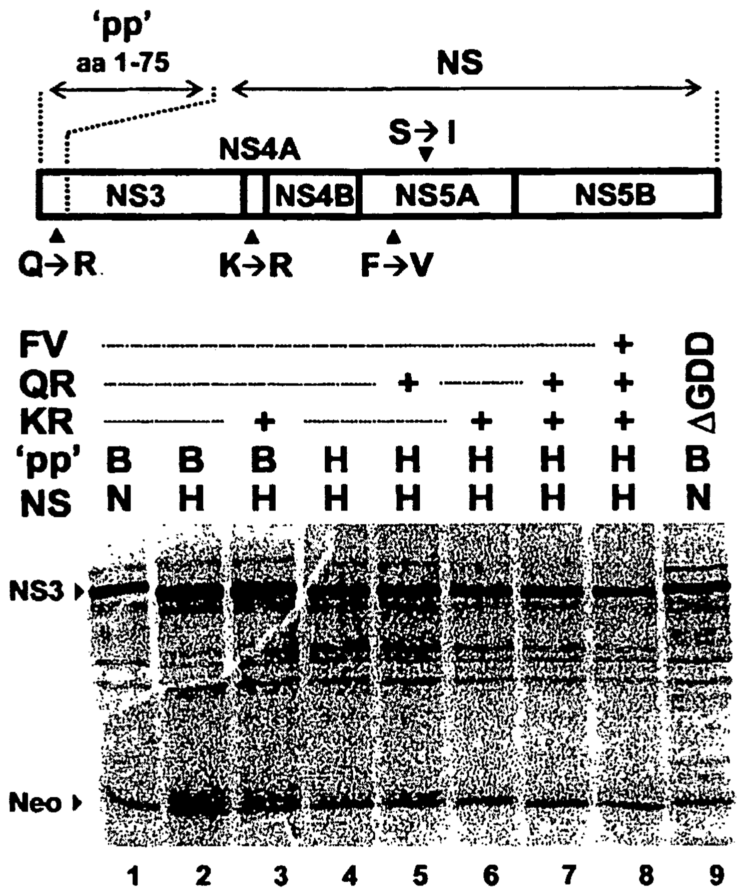


Fig. 4C

	NS Substitutions				
SEAP	3p	3h	4A	5A	5A
-					SI
-			KR		SI
-	QR				SI
+++	QR		KR		SI
-	QR			FV	SI
+			KR	FV	SI
+++	QR		KR	FV	SI
+		PL			SI
+		PL		FV	SI

Fig. 5



'pp'
aa 1-75
NS
S→I
NS4A
NS3
NS4B
NS5A
NS5B
Q→R
K→R
F→V
FV
QR
KR
'pp'
NS
B B B H H H H H B
N H H H H H H H N
ΔGDD
NS3
Neo
1 2 3 4 5 6 7 8 9

Fig. 6A

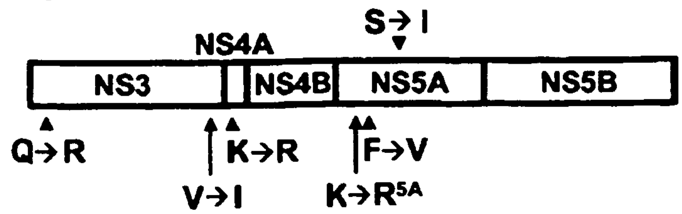

S→I
NS4A
NS3
NS4B
NS5A
NS5B
Q→R
K→R
F→V
V→I
K→R5A

Fig. 6B

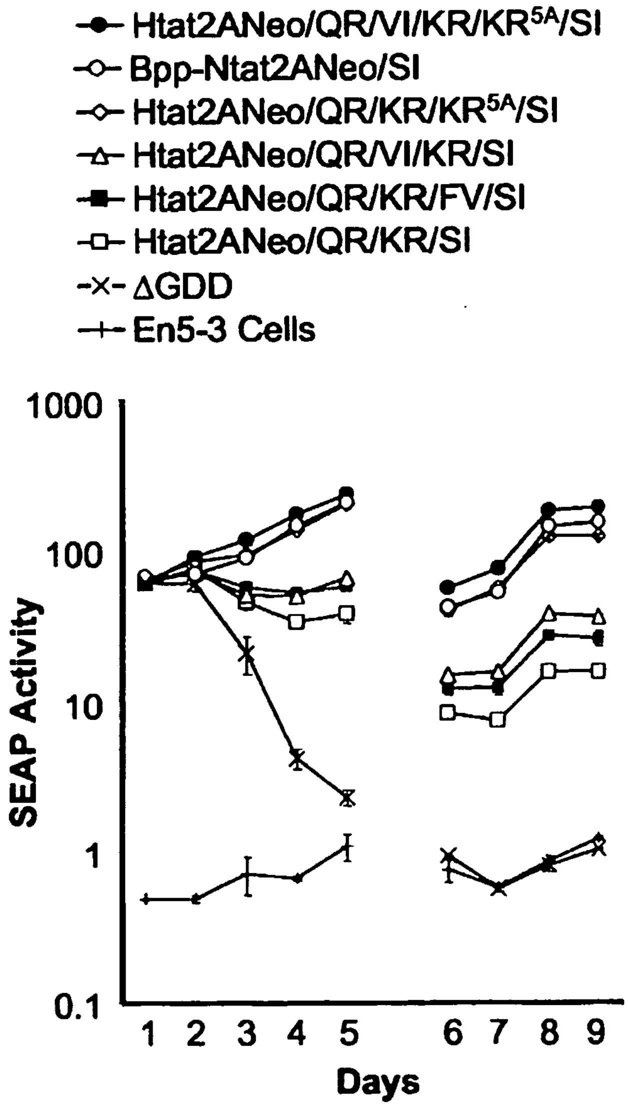
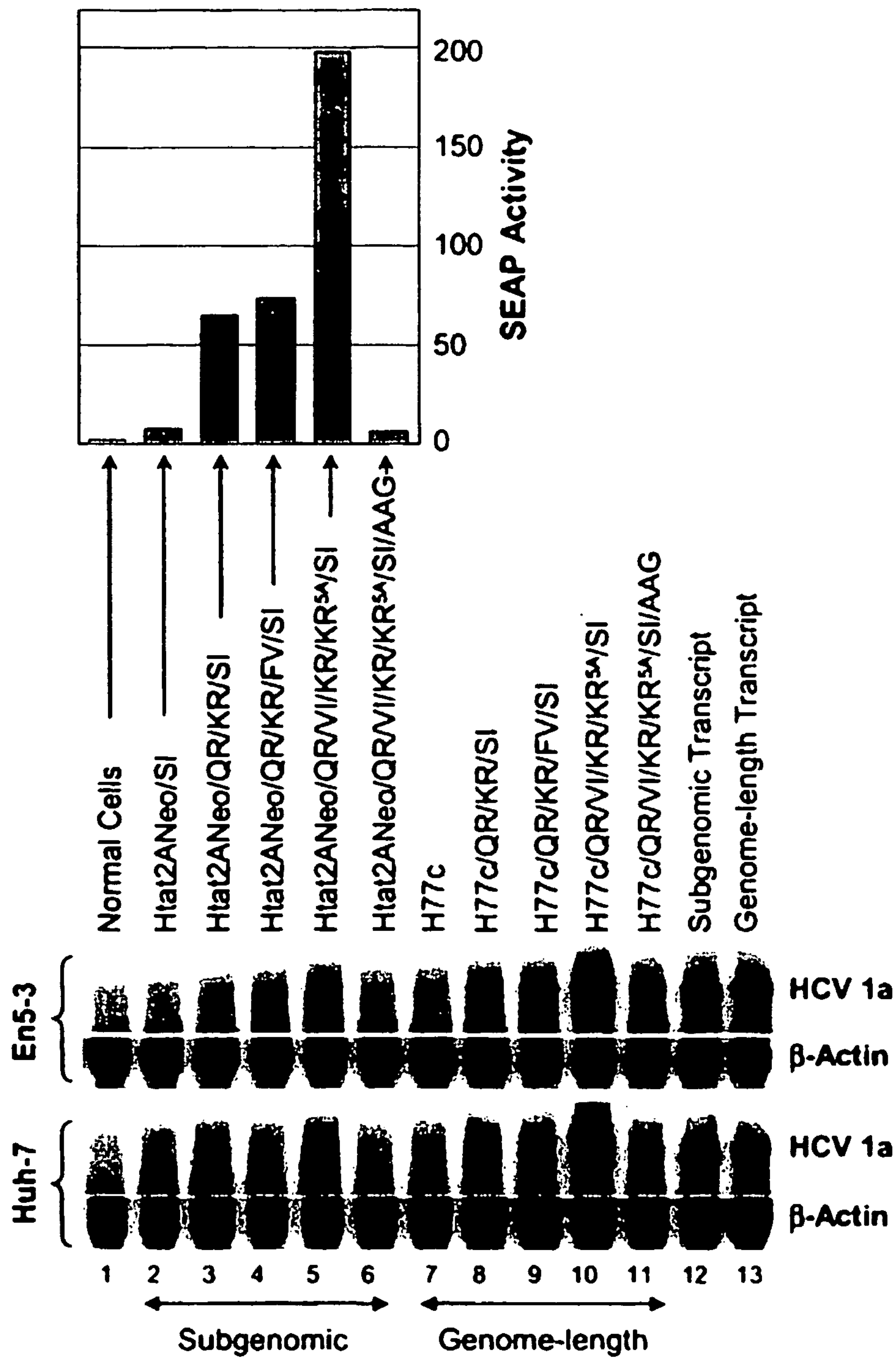

Htat2ANeo/QR/VI/KR/KR5A/SI
Bpp-Ntat2ANeo/SI
Htat2ANeo/QR/KR/KR5A/SI
Htat2ANeo/QR/VI/KR/SI
Htat2ANeo/QR/KR/FV/SI
Htat2ANeo/QR/KR/SI
ΔGDD
En5-3 Cells
1000
100
10
1
0.1
SEAP Activity
1
2
3
4
5
6
7
8
9
Days

Fig. 7



SEAP Activity
0
50
100
150
200
Normal Cells
Htat2ANeo/SI
Htat2ANeo/QR/KR/SI
Htat2ANeo/QR/KR/FV/SI
Htat2ANeo/QR/VI/KR/KR5A/SI
Htat2ANeo/QR/VI/KR/KR5A/SI/AAG
H77c
H77c/QR/KR/SI
H77c/QR/KR/FV/SI
H77c/QR/VI/KR/KR5A/SI
H77c/QR/VI/KR/KR5A/SI/AAG
Subgenomic Transcript
Genome-length Transcript
En5-3
Huh-7
HCV 1a
β-Actin
HCV 1a
β-Actin
1 2 3 4 5 6 7 8 9 10 11 12 13
Subgenomic
Genome-length

Fig. 8A

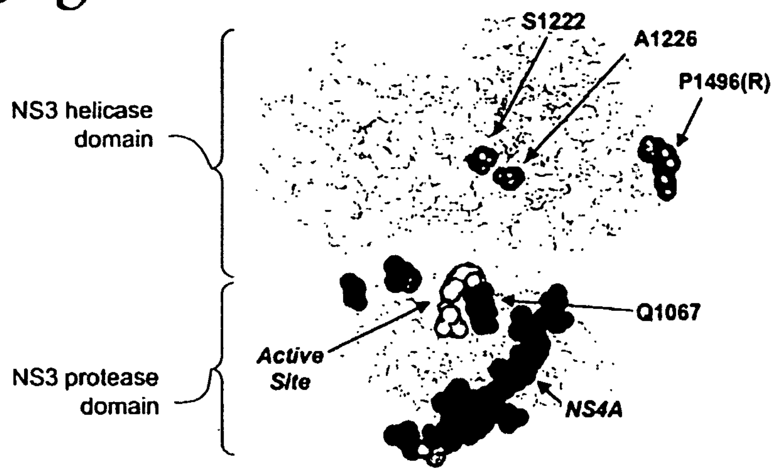


Fig. 8B

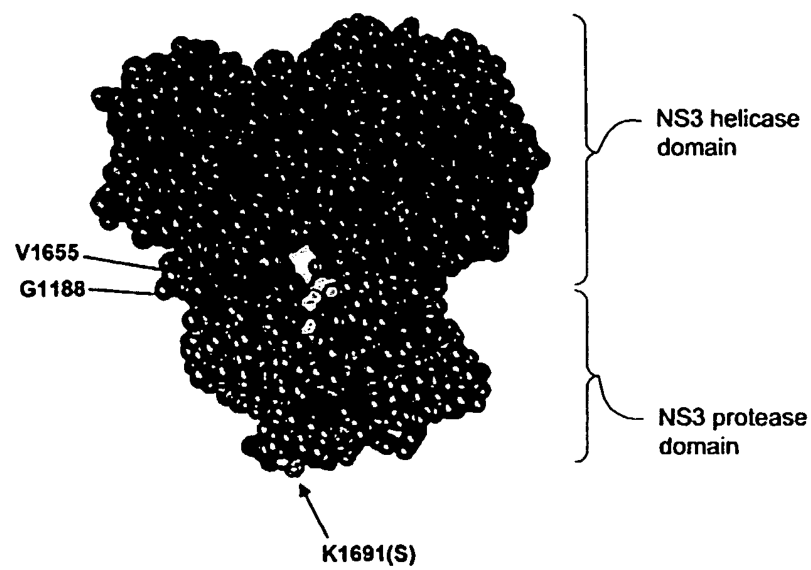


Fig. 8C

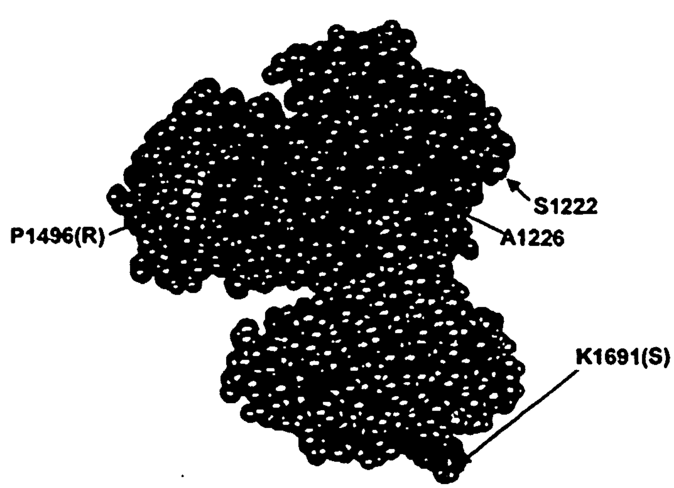


Fig. 9

```
          |  10      |  20      |  30      |  40      |  50      |  60      |  70      |  80
   1 ACCTGGAAAA ACATGGAGCA ATCACAAGTA GCAATACAGC AGCTACCAAT GCTGCTTGTG CCTGGCTAGA AGCACAAGAG 80
  81 GAGGAGGAGG TGGGTTTTCC AGTCACACCT CAGGTACCTT TAAGACCAAT GACTTACAAG GCAGCTGTAG ATCTTAGCCA 160
 161 CTTTTTAAAA GAAAAGGGGG GACTGGAAGG GCTAATTCAC TCCCAAAGAA GACAAGATAT CCTTGATCTG TGGATCTACC 240
 241 ACACACAAGG CTACTTCCCT GATTAGCAGA ACTACACACC AGGGCCAGGG GTCAGATATC CACTGACCTT TGGATGGTGC 320
 321 TACAAGCTAG TACCAGTTGA GCCAGATAAG ATAGAAGAGG CCAATAAAGG AGAGAACACC AGCTTGTTAC ACCCTGTGAG 400
 401 CCTGCATGGG ATGGATGACC CGGAGAGAGA AGTGTTAGAG TGGAGGTTTG ACAGCCGCCT AGCATTTCAT CACGTGGCCC 480
 481 GAGAGCTGCA TCCGGAGTAC TTCAAGAACT GCTGACATCG AGCTTGCTAC AAGGGACTTT CCGCTGGGGA CTTTCCAGGG 560
 561 AGGCGTGGCC TGGGCGGGAC TGGGGAGTGG CGAGCCCTCA GATCCTGCAT ATAAGCAGCT GCTTTTTGCC TGTACTGGGT 640
 641 CTCTCTGGTT AGACCAGATC TGAGCCTGGG AGCTCTCTGG CTAACTAGGG AACCCACTGC TTAAGCCTCA ATAaagcttC 720
 721 TGCATGCTGC TGCTGCTGCT GCTGCTGGGC CTGAGGCTAC AGCTCTCCCT GGGCATCATC CCAGTTGAGG AGGAGAACCC 800
 801 GGACTTCTGG AACCGCGAGG CAGCCGAGGC CCTGGGTGCC GCCAAGAAGC TGCAGCCTGC ACAGACAGCC GCCAAGAACC 880
 881 TCATCATCTT CCTGGGCGAT GGGATGGGGG TGTCTACGGT GACAGCTGCC AGGATCCTAA AAGGGCAGAA GAAGGACAAA 960
 961 CTGGGGCCTG AGATACCCCT GGCCATGGAC CGCTTCCCAT ATGTGGCTCT GTCCAAGACA TACAATGTAG ACAAACATGT 1040
1041 GCCAGACAGT GGAGCCACAG CCACGGCCTA CCTGTGCGGG GTCAAGGGCA ACTTCCAGAC CATTGGCTTG AGTGCAGCCG 1120
1121 CCCGCTTTAA CCAGTGCAAC ACGACACGCG GCAACGAGGT CATCTCCGTG ATGAATCGGG CCAAGAAAGC AGGGAAGTCA 1200
1201 GTGGGAGTGG TAACCACCAC ACGAGTGCAG CACGCCTCGC CAGCCGGCAC CTACGCCCAC ACGGTGAACC GCAACTGGTA 1280
1281 CTCGGACGCC GACGTGCCTG CCTCGGCCCG CCAGGAGGGG TGCCAGGACA TCGCTACGCA GCTCATCTCC AACATGGACA 1360
1361 TTGACGTGAT CCTAGGTGGA GGCCGAAAGT ACATGTTTCC CATGGGAACC CCAGACCCTG AGTACCCAGA TGACTACAGC 1440
1441 CAAGGTGGGA CCAGGCTGGA CGGGAAGAAT CTGGTGCAGG AATGGCTGGC GAAGCGCCAG GGTGCCCGGT ATGTGTGGAA 1520
1521 CCGCACTGAG CTCATGCAGG CTTCCCTGGA CCCGTCTGTG ACCCATCTCA TGGGTCTCTT TGAGCCTGGA GACATGAAAT 1600
1601 ACGAGATCCA CCGAGACTCC ACACTGGACC CCTCCCTGAT GGAGATGACA GAGGCTGCCC TGCGCCTGCT GAGCAGGAAC 1680
1681 CCCCGCGGCT TCTTCCTCTT CGTGGAGGGT GGTCGCATCG ACCATGGTCA TCATGAAAGC AGGGCTTACC GGGCACTGAC 1760
1761 TGAGACGATC ATGTTCGACG ACGCCATTGA GAGGGCGGGC CAGCTCACCA GCGAGGAGGA CACGCTGAGC CTCGTCACTG 1840
1841 CCGACCACTC CCACGTCTTC TCCTTCGGAG GCTACCCCCT GCGAGGGAGC TCCATCTTCG GGCTGGCCCC TGGCAAGGCC 1920
1921 CGGGACAGGA AGGCCTACAC GGTCCTCCTA TACGGAAACG GTCCAGGCTA TGTGCTCAAG GACGGCGCCC GGCCGGATGT 2000
2001 TACCGAGAGC GAGAGCGGGA GCCCCGAGTA TCGGCAGCAG TCAGCAGTGC CCCTGGACGA AGAGACCCAC GCAGGCGAGG 2080
2081 ACGTGGCGGT GTTCGCGCGC GGCCCGCAGG CGCACCTGGT TCACGGCGTG CAGGAGCAGA CCTTCATAGC GCACGTCATG 2160
2161 GCCTTCGCCG CCTGCCTGGA GCCCTACACC GCCTGCGACC TGGCGCCCCC CGCCGGCACC ACCGACGCCG CGCACCCGG  2239
          |  10      |  20      |  30      |  40      |  50      |  60      |  70      |  80
```

Fig. 10A

```
            |   10        |   20        |   30        |   40        |   50        |   60        |   70        |   80        |   90        |  100
  1 TGAAGGTTGG GGTAAACACT CCGGCCTCTT AAGCCATTTC CTGTTTTTTT TTTTTTTTTT TTTTTTTTTT TCTTTTTTTT TTTCTTTCCT TTCCTTCTTT  100
101 TTTTCCTTTC TTTTTCCCTT CTTTAATGGT GGCTCCATCT TAGCCCTAGT CACGGCTAGC TGTGAAAGGT CCGTGAGCCG CATGACTGCA GAGAGTGCTG  200
201 ATACTGGCCT CTCTGCAGAT CATGT
```

Fig. 10B

```
            |   10        |   20        |   30        |   40        |   50        |   60        |   70        |   80        |   90        |  100
  1 GCCAGCCCCC TGATGGGGGC GACACTCCAC CATGAATCAC TCCCCTGTGA GGAACTACTG TCTTCACGCA GAAAGCGTCT AGCCATGGCG TTAGTATGAG 100
101 TGTCGTGCAG CCTCCAGGAC CCCCCCTCCC GGGAGAGCCA TAGTGGTCTG CGGAACCGGT GAGTACACCG GAATTGCCAG GACGACCGGG TCCTTTCTTG 200
201 GATAAACCCG CTCAATGCCT GGAGATTTGG GCGTGCCCCC GCAAGACTGC TAGCCGAGTA GTGTTGGGTC GCGAAAGGCC TTGTGGTACT GCCTGATAGG 300
301 GTGCTTGCGA GTGCCCCGGG AGGTCTCGTA GACCGTGCAC C
```

Fig. 11A-1

```
        |  10       |  20       |  30       |  40       |  50        |  60       |  70       |  80       |  90       |  100
   1 GCCAGCCCCC TGATGGGGGC GACACTCCAC CATGAATCAC TCCCCTGTGA GGAACTACTG TCTTCACGCA GAAAGCGTCT AGCCATGGCG TTAGTATGAG 100
 101 TGTCGTGCAG CCTCCAGGAC CCCCCCTCCC GGGAGAGCCA TAGTGGTCTG CGGAACCGGT GAGTACACCG GAATTGCCAG GACGACCGGG TCCTTTCTTG 200
 201 GATAAACCCG CTCAATGCCT GGAGATTTGG GCGTGCCCCC GCAAGACTGC TAGCCGAGTA GTGTTGGGTC GCGAAAGGCC TTGTGGTACT GCCTGATAGG 300
 301 GTGCTTGCGA GTGCCCCGGG AGGTCTCGTA GACCGTGCAC CATGAGCACG AATCCTAAAC CTCAAAGAAA AACCAAACGT AACACCAACC GTCGCCCACA 400
 401 GGACGTCAAG TTCCCGGGTG GCGGTCAGAT CGTTGGTGGA GTTTACTTGT TGCCGCGCAG GGGCCCTAGA TTGGGTGTGC GCGCGACGAG GAAGACTTCC 500
 501 GAGCGGTCGC AACCTCGAGG TAGACGTCAG CCTATCCCCA AGGCACGTCG GCCCGAGGGC AGGACCTGGG CTCAGCCCGG GTACCCTTGG CCCCTCTATG 600
 601 GCAATGAGGG TTGCGGGTGG GCGGGATGGC TCCTGTCTCC CCGTGGCTCT CGGCCTAGCT GGGGCCCCAC AGACCCCCGG CGTAGGTCGC GCAATTTGGG 700
 701 TAAGGTCATC GATACCCTTA CGTGCGGCTT CGCCGACCTC ATGGGGTACA TACCGCTCGT CGGCGCCCCT CTTGGAGGCG CTGCCAGGGC CCTGGCGCAT 800
 801 GGCGTCCGGG TTCTGGAAGA CGGCGTGAAC TATGCAACAG GGAACCTTCC TGGTTGCTCT TTCTCTATCT TCCTTCTGGC CCTGCTCTCT TGCCTGACTG 900
 901 TGCCCGCTTC AGCCTACCAA GTGCGCAATT CCTCGGGGCT TTACCATGTC ACCAATGATT GCCCTAACTC GAGTATTGTG TACGAGGCGG CCGATGCCAT 1000
1001 CCTGCACACT CCGGGGTGTG TCCCTTGCGT TCGCGAGGGT AACGCCTCGA GGTGTTGGGT GGCGGTGACC CCCACGGTGG CCACCAGGGA CGGCAAACTC 1100
1101 CCCACAACGC AGCTTCGACG TCATATCGAT CTGCTTGTCG GGAGCGCCAC CCTCTGCTCG GCCCTCTACG TGGGGGACCT GTGCGGGTCT GTCTTTCTTG 1200
1201 TTGGTCAACT GTTTACCTTC TCTCCCAGGC GCCACTGGAC GACGCAAGAC TGCAATTGTT CTATCTATCC CGGCCATATA ACGGGTCATC GCATGGCATG 1300
1301 GGATATGATG ATGAACTGGT CCCCTACGGC AGCGTTGGTG GTAGCTCAGC TGCTCCGGAT CCCACAAGCC ATCATGGACA TGATCGCTGG TGCTCACTGG 1400
1401 GGAGTCCTGG CGGGCATAGC GTATTTCTCC ATGGTGGGGA ACTGGGCGAA GGTCCTGGTA GTGCTGCTGC TATTTGCCGG CGTCGACGCG GAAACCCACG 1500
1501 TCACCGGGGG AAATGCCGGC CGCACCACGG CTGGGCTTGT TGGTCTCCTT ACACCAGGCG CCAAGCAGAA CATCCAACTG ATCAACACCA ACGGCAGTTG 1600
1601 GCACATCAAT AGCACGGCCT TGAATTGCAA TGAAAGCCTT AACACCGGCT GGTTAGCAGG GCTCTTCTAT CAACACAAAT TCAACTCTTC AGGCTGTCCT 1700
1701 GAGAGGTTGG CCAGCTGCCG ACGCCTTACC GATTTTGCCC AGGGCTGGGG TCCTATCAGT TATGCCAACG GAAGCGGCCT CGACGAACGC CCCTACTGCT 1800
1801 GGCACTACCC TCCAAGACCT TGTGGCATTG TGCCCGCAAA GAGCGTGTGT GGCCCGGTAT ATTGCTTCAC TCCCAGCCCC GTGGTGGTGG GAACGACCGA 1900
1901 CAGGTCGGGC GCGCCTACCT ACAGCTGGGG TGCAAATGAT ACGGATGTCT TCGTCCTTAA CAACACCAGG CCACCGCTGG GCAATTGGTT CGGTTGTACC 2000
2001 TGGATGAACT CAACTGGATT CACCAAAGTG TGCGGAGCGC CCCCTTGTGT CATCGGAGGG GTGGGCAACA ACACCTTGCT CTGCCCCACT GATTGCTTCC 2100
2101 GCAAACATCC GGAAGCCACA TACTCTCGGT GCGGCTCCGG TCCCTGGATT ACACCCAGGT GCATGGTCGA CTACCCGTAT AGGCTTTGGC ACTATCCTTG 2200
2201 TACCATCAAT TACACCATAT TCAAAGTCAG GATGTACGTG GGAGGGGTCG AGCACAGGCT GGAAGCGGCC TGCAACTGGA CGCGGGGCGA ACGCTGTGAT 2300
2301 CTGGAAGACA GGGACAGGTC CGAGCTCAGC CCGTTGCTGC TGTCCACCAC ACAGTGGCAG GTCCTTCCGT GTTCTTTCAC GACCCTGCCA GCCTTGTCCA 2400
2401 CCGGCCTCAT CCACCTCCAC CAGAACATTG TGGACGTGCA GTACTTGTAC GGGGTAGGGT CAAGCATCGC GTCCTGGGCC ATTAAGTGGG AGTACGTCGT 2500
2501 TCTCCTGTTC CTTCTGCTTG CAGACGCGCG CGTCTGCTCC TGCTTGTGGA TGATGTTACT CATATCCCAA GCGGAGGCGG CTTTGGAGAA CCTCGTAATA 2600
2601 CTCAATGCAG CATCCCTGGC CGGGACGCAC GGTCTTGTGT CCTTCCTCGT GTTCTTCTGC TTTGCGTGGT ATCTGAAGGG TAGGTGGGTG CCCGGAGCGG 2700
2701 TCTACGCCCT CTACGGGATG TGGCCTCTCC TCCTGCTCCT GCTGGCGTTG CCTCAGCGGG CATACGCACT GGACACGGAG GTGGCCGCGT CGTGTGGCGG 2800
2801 CGTTGTTCTT GTCGGGTTAA TGGCGCTGAC TCTGTCGCCA TATTACAAGC GCTATATCAG CTGGTGCATG TGGTGGCTTC AGTATTTTCT GACCAGAGTA 2900
2901 GAAGCGCAAC TGCACGTGTG GGTTCCCCCC CTCAACGTCC GGGGGGGGCG CGATGCCGTC ATCTTACTCA TGTGTGTAGT ACACCCGACC CTGGTATTTG 3000
3001 ACATCACCAA ACTACTCCTG GCCATCTTCG GACCCCTTTG GATTCTTCAA GCCAGTTTGC TTAAAGTCCC CTACTTCGTG CGCGTTCAAG GCCTTCTCCG 3100
3101 GATCTGCGCG CTAGCGCGGA AGATAGCCGG AGGTCATTAC GTGCAAATGG CCATCATCAA GTTAGGGGCG CTTACTGGCA CCTATGTGTA TAACCATCTC 3200
3201 ACCCCTCTTC GAGACTGGGC GCACAACGGC CTGCGAGATC TGGCCGTGGC TGTGGAACCA GTCGTCTTCT CCCGAATGGA GACCAAGCTC ATCACGTGGG 3300
3301 GGGCAGATAC CGCCGCGTGC GGTGACATCA TCAACGGCTT GCCCGTCTCT GCCCGTAGGG GCCAGGAGAT ACTGCTTGGG CCAGCCGACG GAATGGTCTC 3400
3401 CAAGGGGTGG AGGTTGCTGG CGCCCATCAC GGCGTACGCC CAGCAGACGA GAGGCCTCCT AGGGTGTATA ATCACCAGCC TGACTGGCCG GGACAAAAAC 3500
3501 CAAGTGGAGG GTGAGGTCCA GATCGTGTCA ACTGCTACCC AAACCTTCCT GGCAACGTGC ATCAATGGGG TATGCTGGAC TGTCTACCAC GGGGCCGGAA 3600
3601 CGAGGACCAT CGCATCACCC AAGGGTCCTG TCATCCAGAT GTATACCAAT GTGGACCAAG ACCTTGTGGG CTGGCCCGCT CCTCAAGGTT CCCGCTCATT 3700
3701 GACACCCTGT ACCTGCGGCT CCTCGGACCT TTACCTGGTC ACGAGGCACG CCGATGTCAT TCCCGTGCGC CGGCGAGGTG ATAGCAGGGG TAGCCTGCTT 3800
3801 TCGCCCCGGC CCATTTCCTA CTTGAAAGGC TCCTCGGGGG GTCCGCTGTT GTGCCCCGCG GGACACGCCG TGGGCCTATT CAGGGCCGCG GTGTGCACCC 3900
3901 GTGGAGTGGC TAAAGCGGTG GACTTTATCC CTGTGGAGAA CCTAGGGACA ACCATGAGAT CCCCGGTGTT CACGGACAAC TCCTCTCCAC CAGCAGTGCC 4000
4001 CCAGAGCTTC CAGGTGGCCC ACCTGCATGC TCCCACCGGC AGCGGTAAGA GCACCAAGGT CCCGGCTGCG TACGCAGCCC AGGGCTACAA GGTGTTGGTG 4100
```

Fig. 11A-2

```
4101 CTCAACCCCT CTGTTGCTGC AACGCTGGGC TTTGGTGCTT ACATGTCCAA  GGCCCATGGG GTTGATCCTA ATATCAGGAC CGGGGTGAGA ACAATTACCA 4200
4201 CTGGCAGCCC CATCACGTAC TCCACCTACG GCAAGTTCCT TGCCGACGGC  GGGTGCTCAG GAGGTGCTTA TGACATAATA ATTTGTGACG AGTGCCACTC 4300
4301 CACGGATGCC ACATCCATCT TGGGCATCGG CACTGTCCTT GACCAAGCAG  AGACTGCGGG GGCGAGACTG GTTGTGCTCG CCACTGCTAC CCCTCCGGGC 4400
4401 TCCGTCACTG TGTCCCATCC TAACATCGAG GAGGTTGCTC TGTCCACCAC  CGGAGAGATC CCCTTTTACG GCAAGGCTAT CCCCCTCGAG GTGATCAAGG 4500
4501 GGGGAAGACA TCTCATCTTC TGCCACTCAA AGAAGAAGTG CGACGAGCTC  GCCGCGAAGC TGGTCGCATT GGGCATCAAT GCCGTGGCCT ACTACCGCGG 4600
4601 TCTTGACGTG TCTGTCATCC CGACCAGCGG CGATGTTGTC GTCGTGTCGA  CCGATGCTCT CATGACTGGC TTTACCGGCG ACTTCGACTC TGTGATAGAC 4700
4701 TGCAACACGT GTGTCACTCA GACAGTCGAT TTCAGCCTTG ACCCTACCTT  TACCATTGAG ACAACCACGC TCCCCCAGGA TGCTGTCTCC AGGACTCAAC 4800
4801 GCCGGGGCAG GACTGGCAGG GGGAAGCCAG GCATCTATAG ATTTGTGGCA  CCGGGGGAGC GCCCCTCCGG CATGTTCGAC TCGTCCGTCC TCTGTGAGTG 4900
4901 CTATGACGCG GGCTGTGCTT GGTATGAGCT CACGCCCGCC GAGACTACAG  TTAGGCTACG AGCGTACATG AACACCCCGG GGCTTCCCGT GTGCCAGGAC 5000
5001 CATCTTGAAT TTTGGGAGGG CGTCTTTACG GGCCTCACTC ATATAGATGC  CCACTTTTTA TCCCAGACAA AGCAGAGTGG GGAGAACTTT CCTTACCTGG 5100
5101 TAGCGTACCA AGCCACCGTG TGCGCTAGGG CTCAAGCCCC TCCCCCATCG  TGGGACCAGA TGTGGAAGTG TTTGATCCGC CTTAAACCCA CCCTCCATGG 5200
5201 GCCAACACCC CTGCTATACA GACTGGGCGC TGTTCAGAAT GAAGTCACCC  TGACGCACCC AATCACCAAA TACATCATGA CATGCATGTC GGCCGACCTG 5300
5301 GAGGTCGTCA CGAGCACCTG GGTGCTCGTT GGCGGCGTCC TGGCTGCTCT  GGCCGCGTAT TGCCTGTCAA CAGGCTGCGT GGTCATAGTG GGCAGGATCG 5400
5401 TCTTGTCCGG GAAGCCGGCA ATTATACCTG ACAGGGAGGT TCTCTACCAG  GAGTTCGATG AGATGGAAGA GTGCTCTCAG CACTTACCGT ACATCGAGCA 5500
5501 AGGGATGATG CTCGCTGAGC AGTTCAAGCA GAAGGCCCTC GGCCTCCTGC  AGACCGCGTC CCGCCATGCA GAGGTTATCA CCCCTGCTGT CCAGACCAAC 5600
5601 TGGCAGAAAC TCGAGGTCTT TTGGGCGAAG CACATGTGGA ATTTCATCAG  TGGGATACAA TACTTGGCGG GCCTGTCAAC GCTGCCTGGT AACCCCGCCA 5700
5701 TTGCTTCATT GATGGCTTTT ACAGCTGCCG TCACCAGCCC ACTAACCACT  GGCCAAACCC TCCTCTTCAA CATATTGGGG GGGTGGGTGG CTGCCCAGCT 5800
5801 CGCCGCCCCC GGTGCCGCTA CTGCCTTTGT GGGTGCTGGC CTAGCTGGCG  CCGCCATCGG CAGCGTTGGA CTGGGGAAGG TCCTCGTGGA CATTCTTGCA 5900
5901 GGGTATGGCG CGGGCGTGGC GGGAGCTCTT GTAGCATTCA AGATCATGAG  CGGTGAGGTC CCCTCCACGG AGGACCTGGT CAATCTGCTG CCCGCCATCC 6000
6001 TCTCGCCTGG AGCCCTTGTA GTCGGTGTGG TCTGCGCAGC AATACTGCGC  CGGCACGTTG GCCCGGGCGA GGGGGCAGTG CAATGGATGA ACCGGCTAAT 6100
6101 AGCCTTCGCC TCCCGGGGGA ACCATGTTTC CCCCACGCAC TACGTGCCGG  AGAGCGATGC AGCCGCCCGC GTCACTGCCA TACTCAGCAG CCTCACTGTA 6200
6201 ACCCAGCTCC TGAGGCGACT GCATCAGTGG ATAAGCTCGG AGTGTACCAC  TCCATGCTCC GGTTCCTGGC TAAGGGACAT CTGGGACTGG ATATGCGAGG 6300
6301 TGCTGAGCGA CTTTAAGACC TGGCTGAAAG CCAAGCTCAT GCCACAACTG  CCTGGGATTC CCTTTGTGTC CTGCCAGCGC GGGTATAGGG GGGTCTGGCG 6400
6401 AGGAGACGGC ATTATGCACA CTCGCTGCCA CTGTGGAGCT GAGATCACTG  GACATGTCAA AAACGGGACG ATGAGGATCG TCGGTCCTAG GACCTGCAGG 6500
6501 AACATGTGGA GTGGGACGTT CCCCATTAAC GCCTACACCA CGGGCCCCTG  TACTCCCCTT CCTGCGCCGA ACTATAAGTT CGCGCTGTGG AGGGTGTCTG 6600
6601 CAGAGGAATA CGTGGAGATA AGGCGGGTGG GGGACTTCCA CTACGTATCG  GGTATGACTA CTGACAATCT TAAATGCCCG TGCCAGATCC CATCGCCCGA 6700
6701 ATTTTTCACA GAATTGGACG GGGTGCGCCT ACACAGGTTT GCGCCCCCTT  GCAAGCCCTT GCTGCGGGAG GAGGTATCAT TCAGAGTAGG ACTCCACGAG 6800
6801 TACCCGGTGG GGTCGCAATT ACCTTGCGAG CCCGAACCGG ACGTAGCCGT  GTTGACGTCC ATGCTCACTG ATCCCTCCCA TATAACAGCA GAGGCGGCCG 6900
6901 GGAGAAGGTT GGCGAGAGGG TCACCCCCTT CTATGGCCAG CTCCTCGGCT  AGCCAGCTGT CCGCTCCATC TCTCAAGGCA ACTTGCACCG CCAACCATGA 7000
7001 CTCCCCTGAC GCCGAGCTCA TAGAGGCTAA CCTCCTGTGG AGGCAGGAGA  TGGGCGGCAA CATCACCAGG GTTGAGTCAG AGAACAAAGT GGTGATTCTG 7100
7101 GACTCCTTCG ATCCGCTTGT GGCAGAGGAG GATGAGCGGG AGGTCTCCGT  ACCTGCAGAA ATTCTGCGGA AGTCTCGGAG ATTCGCCCGG GCCCTGCCCG 7200
7201 TCTGGGCGCG GCCGGACTAC AACCCCCCGC TAGTAGAGAC GTGGAAAAAG  CCTGACTACG AACCACCTGT GGTCCATGGC TGCCCGCTAC CACCTCCACG 7300
7301 GTCCCCTCCT GTGCCTCCGC CTCGGAAAAA GCGTACGGTG GTCCTCACCG  AATCAACCCT ATCTACTGCC TTGGCCGAGC TTGCCACCAA AAGTTTTGGC 7400
7401 AGCTCCTCAA CTTCCGGCAT TACGGGCGAC AATACGACAA CATCCTCTGA  GCCCGCCCCT TCTGGCTGCC CCCCCGACTC CGACGTTGAG TCCTATTCTT 7500
7501 CCATGCCCCC CCTGGAGGGG GAGCCTGGGG ATCCGGATCT CAGCGACGGG  TCATGGTCGA CGGTCAGTAG TGGGGCCGAC ACGGAAGATG TCGTGTGCTG 7600
7601 CTCAATGTCT TATTCCTGGA CAGGCGCACT CGTCACCCCG TGCGCTGCGG  AAGAACAAAA ACTGCCCATC AACGCACTGA GCAACTCGTT GCTACGCCAT 7700
7701 CACAATCTGG TGTATTCCAC CACTTCACGC AGTGCTTGCC AAAGGCAGAA  GAAAGTCACA TTTGACAGAC TGCAAGTTCT GGACAGCCAT TACCAGGACG 7800
7801 TGCTCAAGGA GGTCAAAGCA GCGGCGTCAA AAGTGAAGGC TAACTTGCTA  TCCGTAGAGG AAGCTTGCAG CCTGACGCCC CCACATTCAG CCAAATCCAA 7900
7901 GTTTGGCTAT GGGGCAAAAG ACGTCCGTTG CCATGCCAGA AAGGCCGTAG  CCCACATCAA CTCCGTGTGG AAAGACCTTC TGGAAGACAG TGTAACACCA 8000
8001 ATAGACACTA CCATCATGGC CAAGAACGAG GTTTTCTGCG TTCAGCCTGA  GAAGGGGGGT CGTAAGCCAG CTCGTCTCAT CGTGTTCCCC GACCTGGGCG 8100
8101 TGCGCGTGTG CGAGAAGATG GCCCTGTACG ACGTGGTTAG CAAGCTCCCC  CTGGCCGTGA TGGGAAGCTC CTACGGATTC CAATACTCAC CAGGACAGCG 8200
8201 GGTTGAATTC CTCGTGCAAG CGTGGAAGTC CAAGAAGACC CCGATGGGGT  TCTCGTATGA TACCCGCTGT TTTGACTCCA CAGTCACTGA GAGCGACATC 8300
```

Fig. 11A-3

```
8301 CGTACGGAGG AGGCAATTTA CCAATGTTGT GACCTGGACC CCCAAGCCCG CGTGGCCATC AAGTCCCTCA CTGAGAGGCT TTATGTTGGG GGCCCTCTTA 8400
8401 CCAATTCAAG GGGGGAAAAC TGCGGCTACC GCAGGTGCCG CGCGAGCGGC GTACTGACAA CTAGCTGTGG TAACACCCTC ACTTGCTACA TCAAGGCCCG 8500
8501 GGCAGCCTGT CGAGCCGCAG GGCTCCAGGA CTGCACCATG CTCGTGTGTG GCGACGACTT AGTCGTTATC TGTGAAAGTG CGGGGGTCCA GGAGGACGCG 8600
8601 GCGAGCCTGA GAGCCTTCAC GGAGGCTATG ACCAGGTACT CCGCCCCCCC CGGGGACCCC CCACAACCAG AATACGACTT GGAGCTTATA ACATCATGCT 8700
8701 CCTCCAACGT GTCAGTCGCC CACGACGGCG CTGGAAAGAG GGTCTACTAC CTTACCCGTG ACCCTACAAC CCCCCTCGCG AGAGCCGCGT GGGAGACAGC 8800
8801 AAGACACACT CCAGTCAATT CCTGGCTAGG CAACATAATC ATGTTTGCCC CCACACTGTG GGCGAGGATG ATACTGATGA CCCATTTCTT TAGCGTCCTC 8900
8901 ATAGCCAGGG ATCAGCTTGA ACAGGCTCTT AACTGTGAGA TCTACGGAGC CTGCTACTCC ATAGAACCAC TGGATCTACC TCCAATCATT CAAAGACTCC 9000
9001 ATGGCCTCAG CGCATTTTCA CTCCACAGTT ACTCTCCAGG TGAAATCAAT AGGGTGGCCG CATGCCTCAG AAAACTTGGG GTCCCGCCCT TGCGAGCTTG 9100
9101 GAGACACCGG GCCCGGAGCG TCCGCGCTAG GCTTCTGTCC AGAGGAGGCA GGGCTGCCAT ATGTGGCAAG TACCTCTTCA ACTGGGCAGT AAGAACAAAG 9200
9201 CTCAAACTCA CTCCAATAGC GGCCGCTGGC CGGCTGGACT TGTCCGGTTG GTTCACGGCT GGCTACAGCG GGGGAGACAT TTATCACAGC GTGTCTCATG 9300
9301 CCCGGCCCCG CTGGTTCTGG TTTTGCCTAC TCCTGCTCGC TGCAGGGGTA GGCATCTACC TCCTCCCCAA CCGATGAAGG TTGGGGTAAA CACTCCGGCC 9400
9401 TCTTAAGCCA TTTCCTGTTT TTTTTTTTTT TTTTTTTTTT TTTTTCTTTT TTTTTTTCTT TCCTTTCCTT CTTTTTTTCC TTTCTTTTTC CCTTCTTTAA 9500
9501 TGGTGGCTCC ATCTTAGCCC TAGTCACGGC TAGCTGTGAA AGGTCCGTGA GCCGCATGAC TGCAGAGAGT GCTGATACTG GCCTCTCTGC AGATCATGT  9599
          | 10       | 20       | 30       | 40       | 50       | 60       | 70       | 80       | 90       | 100
```

Fig. 11B-1

```
342/1                                     372/11                                    402/21
ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT AAC ACC AAC CGT CGC CCA CAG GAC GTC AAG TTC CCG GGT GGC GGT CAG ATC
M   S   T   N   P   K   P   Q   R   K   T   K   R   N   T   N   R   R   P   Q   D   V   K   F   P   G   G   G   Q   I
432/31                                    462/41                                    492/51
GTT GGT GGA GTT TAC TTG TTG CCG CGC AGG GGC CCT AGA TTG GGT GTG CGC GCG ACG AGG AAG ACT TCC GAG CGG TCG CAA CCT CGA GGT
V   G   G   V   Y   L   L   P   R   R   G   P   R   L   G   V   R   A   T   R   K   T   S   E   R   S   Q   P   R   G
522/61                                    552/71                                    582/81
AGA CGT CAG CCT ATC CCC AAG GCA CGT CGG CCC GAG GGC AGG ACC TGG GCT CAG CCC GGG TAC CCT TGG CCC CTC TAT GGC AAT GAG GGT
R   R   Q   P   I   P   K   A   R   R   P   E   G   R   T   W   A   Q   P   G   Y   P   W   P   L   Y   G   N   E   G
612/91                                    642/101                                   672/111
TGC GGG TGG GCG GGA TGG CTC CTG TCT CCC CGT GGC TCT CGG CCT AGC TGG GGC CCC ACA GAC CCC CGG CGT AGG TCG CGC AAT TTG GGT
C   G   W   A   G   W   L   L   S   P   R   G   S   R   P   S   W   G   P   T   D   P   R   R   R   S   R   N   L   G
702/121                                   732/131                                   762/141
AAG GTC ATC GAT ACC CTT ACG TGC GGC TTC GCC GAC CTC ATG GGG TAC ATA CCG CTC GTC GGC GCC CCT CTT GGA GGC GCT GCC AGG GCC
K   V   I   D   T   L   T   C   G   F   A   D   L   M   G   Y   I   P   L   V   G   A   P   L   G   G   A   A   R   A
792/151                                   822/161                                   852/171
CTG GCG CAT GGC GTC CGG GTT CTG GAA GAC GGC GTG AAC TAT GCA ACA GGG AAC CTT CCT GGT TGC TCT TTC TCT ATC TTC CTT CTG GCC
L   A   H   G   V   R   V   L   E   D   G   V   N   Y   A   T   G   N   L   P   G   C   S   F   S   I   F   L   L   A
882/181                                   912/191                                   942/201
CTG CTC TCT TGC CTG ACT GTG CCC GCT TCA GCC TAC CAA GTG CGC AAT TCC TCG GGG CTT TAC CAT GTC ACC AAT GAT TGC CCT AAC TCG
L   L   S   C   L   T   V   P   A   S   A   Y   Q   V   R   N   S   S   G   L   Y   H   V   T   N   D   C   P   N   S
972/211                                   1002/221                                  1032/231
AGT ATT GTG TAC GAG GCG GCC GAT GCC ATC CTG CAC ACT CCG GGG TGT GTC CCT TGC GTT CGC GAG GGT AAC GCC TCG AGG TGT TGG GTG
S   I   V   Y   E   A   A   D   A   I   L   H   T   P   G   C   V   P   C   V   R   E   G   N   A   S   R   C   W   V
1062/241                                  1092/251                                  1122/261
GCG GTG ACC CCC ACG GTG GCC ACC AGG GAC GGC AAA CTC CCC ACA ACG CAG CTT CGA CGT CAT ATC GAT CTG CTT GTC GGG AGC GCC ACC
A   V   T   P   T   V   A   T   R   D   G   K   L   P   T   T   Q   L   R   R   H   I   D   L   L   V   G   S   A   T
1152/271                                  1182/281                                  1212/291
CTC TGC TCG GCC CTC TAC GTG GGG GAC CTG TGC GGG TCT GTC TTT CTT GTT GGT CAA CTG TTT ACC TTC TCT CCC AGG CGC CAC TGG ACG
L   C   S   A   L   Y   V   G   D   L   C   G   S   V   F   L   V   G   Q   L   F   T   F   S   P   R   R   H   W   T
1242/301                                  1272/311                                  1302/321
ACG CAA GAC TGC AAT TGT TCT ATC TAT CCC GGC CAT ATA ACG GGT CAT CGC ATG GCA TGG GAT ATG ATG ATG AAC TGG TCC CCT ACG GCA
T   Q   D   C   N   C   S   I   Y   P   G   H   I   T   G   H   R   M   A   W   D   M   M   M   N   W   S   P   T   A
1332/331                                  1362/341                                  1392/351
GCG TTG GTG GTA GCT CAG CTG CTC CGG ATC CCA CAA GCC ATC ATG GAC ATG ATC GCT GGT GCT CAC TGG GGA GTC CTG GCG GGC ATA GCG
A   L   V   V   A   Q   L   L   R   I   P   Q   A   I   M   D   M   I   A   G   A   H   W   G   V   L   A   G   I   A
1422/361                                  1452/371                                  1482/381
TAT TTC TCC ATG GTG GGG AAC TGG GCG AAG GTC CTG GTA GTG CTG CTG CTA TTT GCC GGC GTC GAC GCG GAA ACC CAC GTC ACC GGG GGA
Y   F   S   M   V   G   N   W   A   K   V   L   V   V   L   L   L   F   A   G   V   D   A   E   T   H   V   T   G   G
1512/391                                  1542/401                                  1572/411
AAT GCC GGC CGC ACC ACG GCT GGG CTT GTT GGT CTC CTT ACA CCA GGC GCC AAG CAG AAC ATC CAA CTG ATC AAC ACC AAC GGC AGT TGG
N   A   G   R   T   T   A   G   L   V   G   L   L   T   P   G   A   K   Q   N   I   Q   L   I   N   T   N   G   S   W
1602/421                                  1632/431                                  1662/441
CAC ATC AAT AGC ACG GCC TTG AAT TGC AAT GAA AGC CTT AAC ACC GGC TGG TTA GCA GGG CTC TTC TAT CAA CAC AAA TTC AAC TCT TCA
H   I   N   S   T   A   L   N   C   N   E   S   L   N   T   G   W   L   A   G   L   F   Y   Q   H   K   F   N   S   S
```

Fig. 11B-2

```
1692/451                                1722/461                                1752/471
GGC TGT CCT GAG AGG TTG GCC AGC TGC CGA CGC CTT ACC GAT TTT GCC CAG GGC TGG GGT CCT ATC AGT TAT GCC AAC GGA AGC GGC CTC
G   C   P   E   R   L   A   S   C   R   R   L   T   D   F   A   Q   G   W   G   P   I   S   Y   A   N   G   S   G   L
1782/481                                1812/491                                1842/501
GAC GAA CGC CCC TAC TGC TGG CAC TAC CCT CCA AGA CCT TGT GGC ATT GTG CCC GCA AAG AGC GTG TGT GGC CCG GTA TAT TGC TTC ACT
D   E   R   P   Y   C   W   H   Y   P   P   R   P   C   G   I   V   P   A   K   S   V   C   G   P   V   Y   C   F   T
1872/511                                1902/521                                1932/531
CCC AGC CCC GTG GTG GTG GGA ACG ACC GAC AGG TCG GGC GCG CCT ACC TAC AGC TGG GGT GCA AAT GAT ACG GAT GTC TTC GTC CTT AAC
P   S   P   V   V   V   G   T   T   D   R   S   G   A   P   T   Y   S   W   G   A   N   D   T   D   V   F   V   L   N
1962/541                                1992/551                                2022/561
AAC ACC AGG CCA CCG CTG GGC AAT TGG TTC GGT TGT ACC TGG ATG AAC TCA ACT GGA TTC ACC AAA GTG TGC GGA GCG CCC CCT TGT GTC
N   T   R   P   P   L   G   N   W   F   G   C   T   W   M   N   S   T   G   F   T   K   V   C   G   A   P   P   C   V
2052/571                                2082/581                                2112/591
ATC GGA GGG GTG GGC AAC AAC ACC TTG CTC TGC CCC ACT GAT TGC TTC CGC AAA CAT CCG GAA GCC ACA TAC TCT CGG TGC GGC TCC GGT
I   G   G   V   G   N   N   T   L   L   C   P   T   D   C   F   R   K   H   P   E   A   T   Y   S   R   C   G   S   G
2142/601                                2172/611                                2202/621
CCC TGG ATT ACA CCC AGG TGC ATG GTC GAC TAC CCG TAT AGG CTT TGG CAC TAT CCT TGT ACC ATC AAT TAC ACC ATA TTC AAA GTC AGG
P   W   I   T   P   R   C   M   V   D   Y   P   Y   R   L   W   H   Y   P   C   T   I   N   Y   T   I   F   K   V   R
2232/631                                2262/641                                2292/651
ATG TAC GTG GGA GGG GTC GAG CAC AGG CTG GAA GCG GCC TGC AAC TGG ACG CGG GGC GAA CGC TGT GAT CTG GAA GAC AGG GAC AGG TCC
M   Y   V   G   G   V   E   H   R   L   E   A   A   C   N   W   T   R   G   E   R   C   D   L   E   D   R   D   R   S
2322/661                                2352/671                                2382/681
GAG CTC AGC CCG TTG CTG CTG TCC ACC ACA CAG TGG CAG GTC CTT CCG TGT TCT TTC ACG ACC CTG CCA GCC TTG TCC ACC GGC CTC ATC
E   L   S   P   L   L   L   S   T   T   Q   W   Q   V   L   P   C   S   F   T   T   L   P   A   L   S   T   G   L   I
2412/691                                2442/701                                2472/711
CAC CTC CAC CAG AAC ATT GTG GAC GTG CAG TAC TTG TAC GGG GTA GGG TCA AGC ATC GCG TCC TGG GCC ATT AAG TGG GAG TAC GTC GTT
H   L   H   Q   N   I   V   D   V   Q   Y   L   Y   G   V   G   S   S   I   A   S   W   A   I   K   W   E   Y   V   V
2502/721                                2532/731                                2562/741
CTC CTG TTC CTT CTG CTT GCA GAC GCG CGC GTC TGC TCC TGC TTG TGG ATG ATG TTA CTC ATA TCC CAA GCG GAG GCG GCT TTG GAG AAC
L   L   F   L   L   L   A   D   A   R   V   C   S   C   L   W   M   M   L   L   I   S   Q   A   E   A   A   L   E   N
2592/751                                2622/761                                2652/771
CTC GTA ATA CTC AAT GCA GCA TCC CTG GCC GGG ACG CAC GGT CTT GTG TCC TTC CTC GTG TTC TTC TGC TTT GCG TGG TAT CTG AAG GGT
L   V   I   L   N   A   A   S   L   A   G   T   H   G   L   V   S   F   L   V   F   F   C   F   A   W   Y   L   K   G
2682/781                                2712/791                                2742/801
AGG TGG GTG CCC GGA GCG GTC TAC GCC CTC TAC GGG ATG TGG CCT CTC CTC CTG CTC CTG CTG GCG TTG CCT CAG CGG GCA TAC GCA CTG
R   W   V   P   G   A   V   Y   A   L   Y   G   M   W   P   L   L   L   L   L   L   A   L   P   Q   R   A   Y   A   L
2772/811                                2802/821                                2832/831
GAC ACG GAG GTG GCC GCG TCG TGT GGC GGC GTT GTT CTT GTC GGG TTA ATG GCG CTG ACT CTG TCG CCA TAT TAC AAG CGC TAT ATC AGC
D   T   E   V   A   A   S   C   G   G   V   V   L   V   G   L   M   A   L   T   L   S   P   Y   Y   K   R   Y   I   S
2862/841                                2892/851                                2922/861
TGG TGC ATG TGG TGG CTT CAG TAT TTT CTG ACC AGA GTA GAA GCG CAA CTG CAC GTG TGG GTT CCC CCC CTC AAC GTC CGG GGG GGG CGC
W   C   M   W   W   L   Q   Y   F   L   T   R   V   E   A   Q   L   H   V   W   V   P   P   L   N   V   R   G   G   R
2952/871                                2982/881                                3012/891
GAT GCC GTC ATC TTA CTC ATG TGT GTA GTA CAC CCG ACC CTG GTA TTT GAC ATC ACC AAA CTA CTC CTG GCC ATC TTC GGA CCC CTT TGG
D   A   V   I   L   L   M   C   V   V   H   P   T   L   V   F   D   I   T   K   L   L   L   A   I   F   G   P   L   W
```

Fig. 11B-3

```
3042/901                                3072/911                                3102/921
ATT CTT CAA GCC AGT TTG CTT AAA GTC CCC TAC TTC GTG CGC GTT CAA GGC CTT CTC CGG ATC TGC GCG CTA GCG CGG AAG ATA GCC GGA
I   L   Q   A   S   L   L   K   V   P   Y   F   V   R   V   Q   G   L   L   R   I   C   A   L   A   R   K   I   A   G
3132/931                                3162/941                                3192/951
GGT CAT TAC GTG CAA ATG GCC ATC ATC AAG TTA GGG GCG CTT ACT GGC ACC TAT GTG TAT AAC CAT CTC ACC CCT CTT CGA GAC TGG GCG
G   H   Y   V   Q   M   A   I   I   K   L   G   A   L   T   G   T   Y   V   Y   N   H   L   T   P   L   R   D   W   A
3222/961                                3252/971                                3282/981
CAC AAC GGC CTG CGA GAT CTG GCC GTG GCT GTG GAA CCA GTC GTC TTC TCC CGA ATG GAG ACC AAG CTC ATC ACG TGG GGG GCA GAT ACC
H   N   G   L   R   D   L   A   V   A   V   E   P   V   V   F   S   R   M   E   T   K   L   I   T   W   G   A   D   T
3312/991                                3342/1001                               3372/1011
GCC GCG TGC GGT GAC ATC ATC AAC GGC TTG CCC GTC TCT GCC CGT AGG GGC CAG GAG ATA CTG CTT GGG CCA GCC GAC GGA ATG GTC TCC
A   A   C   G   D   I   I   N   G   L   P   V   S   A   R   R   G   Q   E   I   L   L   G   P   A   D   G   M   V   S
3402/1021                               3432/1031                               3462/1041
AAG GGG TGG AGG TTG CTG GCG CCC ATC ACG GCG TAC GCC CAG CAG ACG AGA GGC CTC CTA GGG TGT ATA ATC ACC AGC CTG ACT GGC CGG
K   G   W   R   L   L   A   P   I   T   A   Y   A   Q   Q   T   R   G   L   L   G   C   I   I   T   S   L   T   G   R
3492/1051                               3522/1061                               3552/1071
GAC AAA AAC CAA GTG GAG GGT GAG GTC CAG ATC GTG TCA ACT GCT ACC CAA ACC TTC CTG GCA ACG TGC ATC AAT GGG GTA TGC TGG ACT
D   K   N   Q   V   E   G   E   V   Q   I   V   S   T   A   T   Q   T   F   L   A   T   C   I   N   G   V   C   W   T
3582/1081                               3612/1091                               3642/1101
GTC TAC CAC GGG GCC GGA ACG AGG ACC ATC GCA TCA CCC AAG GGT CCT GTC ATC CAG ATG TAT ACC AAT GTG GAC CAA GAC CTT GTG GGC
V   Y   H   G   A   G   T   R   T   I   A   S   P   K   G   P   V   I   Q   M   Y   T   N   V   D   Q   D   L   V   G
3672/1111                               3702/1121                               3732/1131
TGG CCC GCT CCT CAA GGT TCC CGC TCA TTG ACA CCC TGT ACC TGC GGC TCC TCG GAC CTT TAC CTG GTC ACG AGG CAC GCC GAT GTC ATT
W   P   A   P   Q   G   S   R   S   L   T   P   C   T   C   G   S   S   D   L   Y   L   V   T   R   H   A   D   V   I
3762/1141                               3792/1151                               3822/1161
CCC GTG CGC CGG CGA GGT GAT AGC AGG GGT AGC CTG CTT TCG CCC CGG CCC ATT TCC TAC TTG AAA GGC TCC TCG GGG GGT CCG CTG TTG
P   V   R   R   R   G   D   S   R   G   S   L   L   S   P   R   P   I   S   Y   L   K   G   S   S   G   G   P   L   L
3852/1171                               3882/1181                               3912/1191
TGC CCC GCG GGA CAC GCC GTG GGC CTA TTC AGG GCC GCG GTG TGC ACC CGT GGA GTG GCT AAA GCG GTG GAC TTT ATC CCT GTG GAG AAC
C   P   A   G   H   A   V   G   L   F   R   A   A   V   C   T   R   G   V   A   K   A   V   D   F   I   P   V   E   N
3942/1201                               3972/1211                               4002/1221
CTA GGG ACA ACC ATG AGA TCC CCG GTG TTC ACG GAC AAC TCC TCT CCA CCA GCA GTG CCC CAG AGC TTC CAG GTG GCC CAC CTG CAT GCT
L   G   T   T   M   R   S   P   V   F   T   D   N   S   S   P   P   A   V   P   Q   S   F   Q   V   A   H   L   H   A
4032/1231                               4062/1241                               4092/1251
CCC ACC GGC AGC GGT AAG AGC ACC AAG GTC CCG GCT GCG TAC GCA GCC CAG GGC TAC AAG GTG TTG GTG CTC AAC CCC TCT GTT GCT GCA
P   T   G   S   G   K   S   T   K   V   P   A   A   Y   A   A   Q   G   Y   K   V   L   V   L   N   P   S   V   A   A
4122/1261                               4152/1271                               4182/1281
ACG CTG GGC TTT GGT GCT TAC ATG TCC AAG GCC CAT GGG GTT GAT CCT AAT ATC AGG ACC GGG GTG AGA ACA ATT ACC ACT GGC AGC CCC
T   L   G   F   G   A   Y   M   S   K   A   H   G   V   D   P   N   I   R   T   G   V   R   T   I   T   T   G   S   P
4212/1291                               4242/1301                               4272/1311
ATC ACG TAC TCC ACC TAC GGC AAG TTC CTT GCC GAC GGC GGG TGC TCA GGA GGT GCT TAT GAC ATA ATA ATT TGT GAC GAG TGC CAC TCC
I   T   Y   S   T   Y   G   K   F   L   A   D   G   G   C   S   G   G   A   Y   D   I   I   I   C   D   E   C   H   S
4302/1321                               4332/1331                               4362/1341
ACG GAT GCC ACA TCC ATC TTG GGC ATC GGC ACT GTC CTT GAC CAA GCA GAG ACT GCG GGG GCG AGA CTG GTT GTG CTC GCC ACT GCT ACC
T   D   A   T   S   I   L   G   I   G   T   V   L   D   Q   A   E   T   A   G   A   R   L   V   V   L   A   T   A   T
```

Fig. 11B-4

```
4392/1351                                   4422/1361                                   4452/1371
CCT CCG GGC TCC GTC ACT GTG TCC CAT CCT AAC ATC GAG GAG GTT GCT CTG TCC ACC ACC GGA GAG ATC CCC TTT TAC GGC AAG GCT ATC
P   P   G   S   V   T   V   S   H   P   N   I   E   E   V   A   L   S   T   T   G   E   I   P   F   Y   G   K   A   I
4482/1381                                   4512/1391                                   4542/1401
CCC CTC GAG GTG ATC AAG GGG GGA AGA CAT CTC ATC TTC TGC CAC TCA AAG AAG AAG TGC GAC GAG CTC GCC GCG AAG CTG GTC GCA TTG
P   L   E   V   I   K   G   G   R   H   L   I   F   C   H   S   K   K   K   C   D   E   L   A   A   K   L   V   A   L
4572/1411                                   4602/1421                                   4632/1431
GGC ATC AAT GCC GTG GCC TAC TAC CGC GGT CTT GAC GTG TCT GTC ATC CCG ACC AGC GGC GAT GTT GTC GTC GTG TCG ACC GAT GCT CTC
G   I   N   A   V   A   Y   Y   R   G   L   D   V   S   V   I   P   T   S   G   D   V   V   V   V   S   T   D   A   L
4662/1441                                   4692/1451                                   4722/1461
ATG ACT GGC TTT ACC GGC GAC TTC GAC TCT GTG ATA GAC TGC AAC ACG TGT GTC ACT CAG ACA GTC GAT TTC AGC CTT GAC CCT ACC TTT
M   T   G   F   T   G   D   F   D   S   V   I   D   C   N   T   C   V   T   Q   T   V   D   F   S   L   D   P   T   F
4752/1471                                   4782/1481                                   4812/1491
ACC ATT GAG ACA ACC ACG CTC CCC CAG GAT GCT GTC TCC AGG ACT CAA CGC CGG GGC AGG ACT GGC AGG GGG AAG CCA GGC ATC TAT AGA
T   I   E   T   T   T   L   P   Q   D   A   V   S   R   T   Q   R   R   G   R   T   G   R   G   K   P   G   I   Y   R
4842/1501                                   4872/1511                                   4902/1521
TTT GTG GCA CCG GGG GAG CGC CCC TCC GGC ATG TTC GAC TCG TCC GTC CTC TGT GAG TGC TAT GAC GCG GGC TGT GCT TGG TAT GAG CTC
F   V   A   P   G   E   R   P   S   G   M   F   D   S   S   V   L   C   E   C   Y   D   A   G   C   A   W   Y   E   L
4932/1531                                   4962/1541                                   4992/1551
ACG CCC GCC GAG ACT ACA GTT AGG CTA CGA GCG TAC ATG AAC ACC CCG GGG CTT CCC GTG TGC CAG GAC CAT CTT GAA TTT TGG GAG GGC
T   P   A   E   T   T   V   R   L   R   A   Y   M   N   T   P   G   L   P   V   C   Q   D   H   L   E   F   W   E   G
5022/1561                                   5052/1571                                   5082/1581
GTC TTT ACG GGC CTC ACT CAT ATA GAT GCC CAC TTT TTA TCC CAG ACA AAG CAG AGT GGG GAG AAC TTT CCT TAC CTG GTA GCG TAC CAA
V   F   T   G   L   T   H   I   D   A   H   F   L   S   Q   T   K   Q   S   G   E   N   F   P   Y   L   V   A   Y   Q
5112/1591                                   5142/1601                                   5172/1611
GCC ACC GTG TGC GCT AGG GCT CAA GCC CCT CCC CCA TCG TGG GAC CAG ATG TGG AAG TGT TTG ATC CGC CTT AAA CCC ACC CTC CAT GGG
A   T   V   C   A   R   A   Q   A   P   P   P   S   W   D   Q   M   W   K   C   L   I   R   L   K   P   T   L   H   G
5202/1621                                   5232/1631                                   5262/1641
CCA ACA CCC CTG CTA TAC AGA CTG GGC GCT GTT CAG AAT GAA GTC ACC CTG ACG CAC CCA ATC ACC AAA TAC ATC ATG ACA TGC ATG TCG
P   T   P   L   L   Y   R   L   G   A   V   Q   N   E   V   T   L   T   H   P   I   T   K   Y   I   M   T   C   M   S
5292/1651                                   5322/1661                                   5352/1671
GCC GAC CTG GAG GTC GTC ACG AGC ACC TGG GTG CTC GTT GGC GGC GTC CTG GCT GCT CTG GCC GCG TAT TGC CTG TCA ACA GGC TGC GTG
A   D   L   E   V   V   T   S   T   W   V   L   V   G   G   V   L   A   A   L   A   A   Y   C   L   S   T   G   C   V
5382/1681                                   5412/1691                                   5442/1701
GTC ATA GTG GGC AGG ATC GTC TTG TCC GGG AAG CCG GCA ATT ATA CCT GAC AGG GAG GTT CTC TAC CAG GAG TTC GAT GAG ATG GAA GAG
V   I   V   G   R   I   V   L   S   G   K   P   A   I   I   P   D   R   E   V   L   Y   Q   E   F   D   E   M   E   E
5472/1711                                   5502/1721                                   5532/1731
TGC TCT CAG CAC TTA CCG TAC ATC GAG CAA GGG ATG ATG CTC GCT GAG CAG TTC AAG CAG AAG GCC CTC GGC CTC CTG CAG ACC GCG TCC
C   S   Q   H   L   P   Y   I   E   Q   G   M   M   L   A   E   Q   F   K   Q   K   A   L   G   L   L   Q   T   A   S
5562/1741                                   5592/1751                                   5622/1761
CGC CAT GCA GAG GTT ATC ACC CCT GCT GTC CAG ACC AAC TGG CAG AAA CTC GAG GTC TTT TGG GCG AAG CAC ATG TGG AAT TTC ATC AGT
R   H   A   E   V   I   T   P   A   V   Q   T   N   W   Q   K   L   E   V   F   W   A   K   H   M   W   N   F   I   S
5652/1771                                   5682/1781                                   5712/1791
GGG ATA CAA TAC TTG GCG GGC CTG TCA ACG CTG CCT GGT AAC CCC GCC ATT GCT TCA TTG ATG GCT TTT ACA GCT GCC GTC ACC AGC CCA
G   I   Q   Y   L   A   G   L   S   T   L   P   G   N   P   A   I   A   S   L   M   A   F   T   A   A   V   T   S   P
```

Fig. 11B-5

```
5742/1801                                                   5772/1811                                                   5802/1821
CTA ACC ACT GGC CAA ACC CTC CTC TTC AAC ATA TTG GGG GGG TGG GTG GCT GCC CAG CTC GCC GCC CCC GGT GCC GCT ACT GCC TTT GTG
L   T   T   G   Q   T   L   L   F   N   I   L   G   G   W   V   A   A   Q   L   A   A   P   G   A   A   T   A   F   V
5832/1831                                                   5862/1841                                                   5892/1851
GGT GCT GGC CTA GCT GGC GCC GCC ATC GGC AGC GTT GGA CTG GGG AAG GTC CTC GTG GAC ATT CTT GCA GGG TAT GGC GCG GGC GTG GCG
G   A   G   L   A   G   A   A   I   G   S   V   G   L   G   K   V   L   V   D   I   L   A   G   Y   G   A   G   V   A
5922/1861                                                   5952/1871                                                   5982/1881
GGA GCT CTT GTA GCA TTC AAG ATC ATG AGC GGT GAG GTC CCC TCC ACG GAG GAC CTG GTC AAT CTG CTG CCC GCC ATC CTC TCG CCT GGA
G   A   L   V   A   F   K   I   M   S   G   E   V   P   S   T   E   D   L   V   N   L   L   P   A   I   L   S   P   G
6012/1891                                                   6042/1901                                                   6072/1911
GCC CTT GTA GTC GGT GTG GTC TGC GCA GCA ATA CTG CGC CGG CAC GTT GGC CCG GGC GAG GGG GCA GTG CAA TGG ATG AAC CGG CTA ATA
A   L   V   V   G   V   V   C   A   A   I   L   R   R   H   V   G   P   G   E   G   A   V   Q   W   M   N   R   L   I
6102/1921                                                   6132/1931                                                   6162/1941
GCC TTC GCC TCC CGG GGG AAC CAT GTT TCC CCC ACG CAC TAC GTG CCG GAG AGC GAT GCA GCC GCC CGC GTC ACT GCC ATA CTC AGC AGC
A   F   A   S   R   G   N   H   V   S   P   T   H   Y   V   P   E   S   D   A   A   A   R   V   T   A   I   L   S   S
6192/1951                                                   6222/1961                                                   6252/1971
CTC ACT GTA ACC CAG CTC CTG AGG CGA CTG CAT CAG TGG ATA AGC TCG GAG TGT ACC ACT CCA TGC TCC GGT TCC TGG CTA AGG GAC ATC
L   T   V   T   Q   L   L   R   R   L   H   Q   W   I   S   S   E   C   T   T   P   C   S   G   S   W   L   R   D   I
6282/1981                                                   6312/1991                                                   6342/2001
TGG GAC TGG ATA TGC GAG GTG CTG AGC GAC TTT AAG ACC TGG CTG AAA GCC AAG CTC ATG CCA CAA CTG CCT GGG ATT CCC TTT GTG TCC
W   D   W   I   C   E   V   L   S   D   F   K   T   W   L   K   A   K   L   M   P   Q   L   P   G   I   P   F   V   S
6372/2011                                                   6402/2021                                                   6432/2031
TGC CAG CGC GGG TAT AGG GGG GTC TGG CGA GGA GAC GGC ATT ATG CAC ACT CGC TGC CAC TGT GGA GCT GAG ATC ACT GGA CAT GTC AAA
C   Q   R   G   Y   R   G   V   W   R   G   D   G   I   M   H   T   R   C   H   C   G   A   E   I   T   G   H   V   K
6462/2041                                                   6492/2051                                                   6522/2061
AAC GGG ACG ATG AGG ATC GTC GGT CCT AGG ACC TGC AGG AAC ATG TGG AGT GGG ACG TTC CCC ATT AAC GCC TAC ACC ACG GGC CCC TGT
N   G   T   M   R   I   V   G   P   R   T   C   R   N   M   W   S   G   T   F   P   I   N   A   Y   T   T   G   P   C
6552/2071                                                   6582/2081                                                   6612/2091
ACT CCC CTT CCT GCG CCG AAC TAT AAG TTC GCG CTG TGG AGG GTG TCT GCA GAG GAA TAC GTG GAG ATA AGG CGG GTG GGG GAC TTC CAC
T   P   L   P   A   P   N   Y   K   F   A   L   W   R   V   S   A   E   E   Y   V   E   I   R   R   V   G   D   F   H
6642/2101                                                   6672/2111                                                   6702/2121
TAC GTA TCG GGT ATG ACT ACT GAC AAT CTT AAA TGC CCG TGC CAG ATC CCA TCG CCC GAA TTT TTC ACA GAA TTG GAC GGG GTG CGC CTA
Y   V   S   G   M   T   T   D   N   L   K   C   P   C   Q   I   P   S   P   E   F   F   T   E   L   D   G   V   R   L
6732/2131                                                   6762/2141                                                   6792/2151
CAC AGG TTT GCG CCC CCT TGC AAG CCC TTG CTG CGG GAG GAG GTA TCA TTC AGA GTA GGA CTC CAC GAG TAC CCG GTG GGG TCG CAA TTA
H   R   F   A   P   P   C   K   P   L   L   R   E   E   V   S   F   R   V   G   L   H   E   Y   P   V   G   S   Q   L
6822/2161                                                   6852/2171                                                   6882/2181
CCT TGC GAG CCC GAA CCG GAC GTA GCC GTG TTG ACG TCC ATG CTC ACT GAT CCC TCC CAT ATA ACA GCA GAG GCG GCC GGG AGA AGG TTG
P   C   E   P   E   P   D   V   A   V   L   T   S   M   L   T   D   P   S   H   I   T   A   E   A   A   G   R   R   L
6912/2191                                                   6942/2201                                                   6972/2211
GCG AGA GGG TCA CCC CCT TCT ATG GCC AGC TCC TCG GCT AGC CAG CTG TCC GCT CCA TCT CTC AAG GCA ACT TGC ACC GCC AAC CAT GAC
A   R   G   S   P   P   S   M   A   S   S   S   A   S   Q   L   S   A   P   S   L   K   A   T   C   T   A   N   H   D
7002/2221                                                   7032/2231                                                   7062/2241
TCC CCT GAC GCC GAG CTC ATA GAG GCT AAC CTC CTG TGG AGG CAG GAG ATG GGC GGC AAC ATC ACC AGG GTT GAG TCA GAG AAC AAA GTG
S   P   D   A   E   L   I   E   A   N   L   L   W   R   Q   E   M   G   G   N   I   T   R   V   E   S   E   N   K   V
```

Fig. 11B-6

```
7092/2251                               7122/2261                               7152/2271
GTG ATT CTG GAC TCC TTC GAT CCG CTT GTG GCA GAG GAG GAT GAG CGG GAG GTC TCC GTA CCT GCA GAA ATT CTG CGG AAG TCT CGG AGA
V   I   L   D   S   F   D   P   L   V   A   E   E   D   E   R   E   V   S   V   P   A   E   I   L   R   K   S   R   R
7182/2281                               7212/2291                               7242/2301
TTC GCC CGG GCC CTG CCC GTC TGG GCG CGG CCG GAC TAC AAC CCC CCG CTA GTA GAG ACG TGG AAA AAG CCT GAC TAC GAA CCA CCT GTG
F   A   R   A   L   P   V   W   A   R   P   D   Y   N   P   P   L   V   E   T   W   K   K   P   D   Y   E   P   P   V
7272/2311                               7302/2321                               7332/2331
GTC CAT GGC TGC CCG CTA CCA CCT CCA CGG TCC CCT CCT GTG CCT CCG CCT CGG AAA AAG CGT ACG GTG GTC CTC ACC GAA TCA ACC CTA
V   H   G   C   P   L   P   P   P   R   S   P   P   V   P   P   P   R   K   K   R   T   V   V   L   T   E   S   T   L
7362/2341                               7392/2351                               7422/2361
TCT ACT GCC TTG GCC GAG CTT GCC ACC AAA AGT TTT GGC AGC TCC TCA ACT TCC GGC ATT ACG GGC GAC AAT ACG ACA ACA TCC TCT GAG
S   T   A   L   A   E   L   A   T   K   S   F   G   S   S   S   T   S   G   I   T   G   D   N   T   T   T   S   S   E
7452/2371                               7482/2381                               7512/2391
CCC GCC CCT TCT GGC TGC CCC CCC GAC TCC GAC GTT GAG TCC TAT TCT TCC ATG CCC CCC CTG GAG GGG GAG CCT GGG GAT CCG GAT CTC
P   A   P   S   G   C   P   P   D   S   D   V   E   S   Y   S   S   M   P   P   L   E   G   E   P   G   D   P   D   L
7542/2401                               7572/2411                               7602/2421
AGC GAC GGG TCA TGG TCG ACG GTC AGT AGT GGG GCC GAC ACG GAA GAT GTC GTG TGC TGC TCA ATG TCT TAT TCC TGG ACA GGC GCA CTC
S   D   G   S   W   S   T   V   S   S   G   A   D   T   E   D   V   V   C   C   S   M   S   Y   S   W   T   G   A   L
7632/2431                               7662/2441                               7692/2451
GTC ACC CCG TGC GCT GCG GAA GAA CAA AAA CTG CCC ATC AAC GCA CTG AGC AAC TCG TTG CTA CGC CAT CAC AAT CTG GTG TAT TCC ACC
V   T   P   C   A   A   E   E   Q   K   L   P   I   N   A   L   S   N   S   L   L   R   H   H   N   L   V   Y   S   T
7722/2461                               7752/2471                               7782/2481
ACT TCA CGC AGT GCT TGC CAA AGG CAG AAG AAA GTC ACA TTT GAC AGA CTG CAA GTT CTG GAC AGC CAT TAC CAG GAC GTG CTC AAG GAG
T   S   R   S   A   C   Q   R   Q   K   K   V   T   F   D   R   L   Q   V   L   D   S   H   Y   Q   D   V   L   K   E
7812/2491                               7842/2501                               7872/2511
GTC AAA GCA GCG GCG TCA AAA GTG AAG GCT AAC TTG CTA TCC GTA GAG GAA GCT TGC AGC CTG ACG CCC CCA CAT TCA GCC AAA TCC AAG
V   K   A   A   A   S   K   V   K   A   N   L   L   S   V   E   E   A   C   S   L   T   P   P   H   S   A   K   S   K
7902/2521                               7932/2531                               7962/2541
TTT GGC TAT GGG GCA AAA GAC GTC CGT TGC CAT GCC AGA AAG GCC GTA GCC CAC ATC AAC TCC GTG TGG AAA GAC CTT CTG GAA GAC AGT
F   G   Y   G   A   K   D   V   R   C   H   A   R   K   A   V   A   H   I   N   S   V   W   K   D   L   L   E   D   S
7992/2551                               8022/2561                               8052/2571
GTA ACA CCA ATA GAC ACT ACC ATC ATG GCC AAG AAC GAG GTT TTC TGC GTT CAG CCT GAG AAG GGG GGT CGT AAG CCA GCT CGT CTC ATC
V   T   P   I   D   T   T   I   M   A   K   N   E   V   F   C   V   Q   P   E   K   G   G   R   K   P   A   R   L   I
8082/2581                               8112/2591                               8142/2601
GTG TTC CCC GAC CTG GGC GTG CGC GTG TGC GAG AAG ATG GCC CTG TAC GAC GTG GTT AGC AAG CTC CCC CTG GCC GTG ATG GGA AGC TCC
V   F   P   D   L   G   V   R   V   C   E   K   M   A   L   Y   D   V   V   S   K   L   P   L   A   V   M   G   S   S
8172/2611                               8202/2621                               8232/2631
TAC GGA TTC CAA TAC TCA CCA GGA CAG CGG GTT GAA TTC CTC GTG CAA GCG TGG AAG TCC AAG AAG ACC CCG ATG GGG TTC TCG TAT GAT
Y   G   F   Q   Y   S   P   G   Q   R   V   E   F   L   V   Q   A   W   K   S   K   K   T   P   M   G   F   S   Y   D
8262/2641                               8292/2651                               8322/2661
ACC CGC TGT TTT GAC TCC ACA GTC ACT GAG AGC GAC ATC CGT ACG GAG GAG GCA ATT TAC CAA TGT TGT GAC CTG GAC CCC CAA GCC CGC
T   R   C   F   D   S   T   V   T   E   S   D   I   R   T   E   E   A   I   Y   Q   C   C   D   L   D   P   Q   A   R
8352/2671                               8382/2681                               8412/2691
GTG GCC ATC AAG TCC CTC ACT GAG AGG CTT TAT GTT GGG GGC CCT CTT ACC AAT TCA AGG GGG GAA AAC TGC GGC TAC CGC AGG TGC CGC
V   A   I   K   S   L   T   E   R   L   Y   V   G   G   P   L   T   N   S   R   G   E   N   C   G   Y   R   R   C   R
```

Fig. 11B-7

```
8442/2701                                                   8472/2711                                                   8502/2721
GCG AGC GGC GTA CTG ACA ACT AGC TGT GGT AAC ACC CTC ACT TGC TAC ATC AAG GCC CGG GCA GCC TGT CGA GCC GCA GGG CTC CAG GAC
A   S   G   V   L   T   T   S   C   G   N   T   L   T   C   Y   I   K   A   R   A   A   C   R   A   A   G   L   Q   D
8532/2731                                                   8562/2741                                                   8592/2751
TGC ACC ATG CTC GTG TGT GGC GAC GAC TTA GTC GTT ATC TGT GAA AGT GCG GGG GTC CAG GAG GAC GCG GCG AGC CTG AGA GCC TTC ACG
C   T   M   L   V   C   G   D   D   L   V   V   I   C   E   S   A   G   V   Q   E   D   A   A   S   L   R   A   F   T
8622/2761                                                   8652/2771                                                   8682/2781
GAG GCT ATG ACC AGG TAC TCC GCC CCC CCC GGG GAC CCC CCA CAA CCA GAA TAC GAC TTG GAG CTT ATA ACA TCA TGC TCC TCC AAC GTG
E   A   M   T   R   Y   S   A   P   P   G   D   P   P   Q   P   E   Y   D   L   E   L   I   T   S   C   S   S   N   V
8712/2791                                                   8742/2801                                                   8772/2811
TCA GTC GCC CAC GAC GGC GCT GGA AAG AGG GTC TAC TAC CTT ACC CGT GAC CCT ACA ACC CCC CTC GCG AGA GCC GCG TGG GAG ACA GCA
S   V   A   H   D   G   A   G   K   R   V   Y   Y   L   T   R   D   P   T   T   P   L   A   R   A   A   W   E   T   A
8802/2821                                                   8832/2831                                                   8862/2841
AGA CAC ACT CCA GTC AAT TCC TGG CTA GGC AAC ATA ATC ATG TTT GCC CCC ACA CTG TGG GCG AGG ATG ATA CTG ATG ACC CAT TTC TTT
R   H   T   P   V   N   S   W   L   G   N   I   I   M   F   A   P   T   L   W   A   R   M   I   L   M   T   H   F   F
8892/2851                                                   8922/2861                                                   8952/2871
AGC GTC CTC ATA GCC AGG GAT CAG CTT GAA CAG GCT CTT AAC TGT GAG ATC TAC GGA GCC TGC TAC TCC ATA GAA CCA CTG GAT CTA CCT
S   V   L   I   A   R   D   Q   L   E   Q   A   L   N   C   E   I   Y   G   A   C   Y   S   I   E   P   L   D   L   P
8982/2881                                                   9012/2891                                                   9042/2901
CCA ATC ATT CAA AGA CTC CAT GGC CTC AGC GCA TTT TCA CTC CAC AGT TAC TCT CCA GGT GAA ATC AAT AGG GTG GCC GCA TGC CTC AGA
P   I   I   Q   R   L   H   G   L   S   A   F   S   L   H   S   Y   S   P   G   E   I   N   R   V   A   A   C   L   R
9072/2911                                                   9102/2921                                                   9132/2931
AAA CTT GGG GTC CCG CCC TTG CGA GCT TGG AGA CAC CGG GCC CGG AGC GTC CGC GCT AGG CTT CTG TCC AGA GGA GGC AGG GCT GCC ATA
K   L   G   V   P   P   L   R   A   W   R   H   R   A   R   S   V   R   A   R   L   L   S   R   G   G   R   A   A   I
9162/2941                                                   9192/2951                                                   9222/2961
TGT GGC AAG TAC CTC TTC AAC TGG GCA GTA AGA ACA AAG CTC AAA CTC ACT CCA ATA GCG GCC GCT GGC CGG CTG GAC TTG TCC GGT TGG
C   G   K   Y   L   F   N   W   A   V   R   T   K   L   K   L   T   P   I   A   A   A   G   R   L   D   L   S   G   W
9252/2971                                                   9282/2981                                                   9312/2991
TTC ACG GCT GGC TAC AGC GGG GGA GAC ATT TAT CAC AGC GTG TCT CAT GCC CGG CCC CGC TGG TTC TGG TTT TGC CTA CTC CTG CTC GCT
F   T   A   G   Y   S   G   G   D   I   Y   H   S   V   S   H   A   R   P   R   W   F   W   F   C   L   L   L   L   A
9342/3001                                                   9372/3011                                                   9402/3021
GCA GGG GTA GGC ATC TAC CTC CTC CCC AAC CGA TGA
A   G   V   G   I   Y   L   L   P   N   R   *
```

Fig. 12A-1

```
         | 10       | 20       | 30       | 40       | 50        | 60       | 70       | 80       | 90       | 100
   1 GCCAGCCCCC TGATGGGGGC GACACTCCAC CATGAATCAC TCCCCTGTGA  GGAACTACTG TCTTCACGCA GAAAGCGTCT AGCCATGGCG TTAGTATGAG 100
 101 TGTCGTGCAG CCTCCAGGAC CCCCCCTCCC GGGAGAGCCA TAGTGGTCTG  CGGAACCGGT GAGTACACCG GAATTGCCAG GACGACCGGG TCCTTTCTTG 200
 201 GATAAACCCG CTCAATGCCT GGAGATTTGG GCGTGCCCCC GCAAGACTGC  TAGCCGAGTA GTGTTGGGTC GCGAAAGGCC TTGTGGTACT GCCTGATAGG 300
 301 GTGCTTGCGA GTGCCCCGGG AGGTCTCGTA GACCGTGCAC CATGGAGCCA  GTAGATCCTA GACTAGAGCC CTGGAAGCAT CCAGGAAGTC AGCCTAAAAC 400
 401 TGCTTGTACC AATTGCTATT GTAAAAAGTG TTGCTTTCAT TGCCAAGTTT  GTTTCATAAC AAAAGCCTTA GGCATCTCCT ATGGCAGGAA GAAGCGGAGA 500
 501 CAGCGACGAA GACCTCCTCA AGGCAGTCAG ACTCATCAAG TTTCTCTATC  AAAGCAACCC ACCTCCCAAT CCCGAGGGGA CCCGACAGGC CCGAAGGAAG 600
 601 AATTCGACCT TCTTAAGCTT GCGGGAGACG TCGAGTCCAA CCCTGGGCCC  GGATCTGTTA ACATGATTGA ACAAGATGGA TTGCACGCAG GTTCTCCGGC 700
 701 CGCTTGGGTG GAGAGGCTAT TCGGCTATGA CTGGGCACAA CAGACAATCG  GCTGCTCTGA TGCCGCCGTG TTCCGGCTGT CAGCGCAGGG GCGCCCGGTT 800
 801 CTTTTTGTCA AGACCGACCT GTCCGGTGCC CTGAATGAAC TGCAGGACGA  GGCAGCGCGG CTATCGTGGC TGGCCACGAC GGGCGTTCCT TGCGCAGCTG 900
 901 TGCTCGACGT TGTCACTGAA GCGGGAAGGG ACTGGCTGCT ATTGGGCGAA  GTGCCGGGGC AGGATCTCCT GTCATCTCAC CTTGCTCCTG CCGAGAAAGT 1000
1001 ATCCATCATG GCTGATGCAA TGCGGCGGCT GCATACGCTT GATCCGGCTA  CCTGCCCATT CGACCACCAA GCGAAACATC GCATCGAGCG AGCACGTACT 1100
1101 CGGATGGAAG CCGGTCTTGT CGATCAGGAT GATCTGGACG AAGAGCATCA  GGGGCTCGCG CCAGCCGAAC TGTTCGCCAG GCTCAAGGCG CGCATGCCCG 1200
1201 ACGGCGAGGA TCTCGTCGTG ACCCATGGCG ATGCCTGCTT GCCGAATATC  ATGGTGGAAA ATGGCCGCTT TTCTGGATTC ATCGACTGTG GCCGGCTGGG 1300
1301 TGTGGCGGAC CGCTATCAGG ACATAGCGTT GGCTACCCGT GATATTGCTG  AAGAGCTTGG CGGCGAATGG GCTGACCGCT TCCTCGTGCT TTACGGTATC 1400
1401 GCCGCTCCCG ATTCGCAGCG CATCGCCTTC TATCGCCTTC TTGACGAGTT  CTTCTGAGTT TAAACAGACC ACAACGGTTT CCCTCTAGCG GGATCAATTC 1500
1501 CGCCCCTCTC CCTCCCCCCC CCCTAACGTT ACTGGCCGAA GCCGCTTGGA  ATAAGGCCGG TGTGCGTTTG TCTATATGTT ATTTTCCACC ATATTGCCGT 1600
1601 CTTTTGGCAA TGTGAGGGCC CGGAAACCTG GCCCTGTCTT CTTGACGAGC  ATTCCTAGGG GTCTTTCCCC TCTCGCCAAA GGAATGCAAG GTCTGTTGAA 1700
1701 TGTCGTGAAG GAAGCAGTTC CTCTGGAAGC TTCTTGAAGA CAAACAACGT  CTGTAGCGAC CCTTTGCAGG CAGCGGAACC CCCCACCTGG CGACAGGTGC 1800
1801 CTCTGCGGCC AAAAGCCACG TGTATAAGAT ACACCTGCAA AGGCGGCACA  ACCCCAGTGC CACGTTGTGA GTTGGATAGT TGTGGAAAGA GTCAAATGGC 1900
1901 TCTCCTCAAG CGTATTCAAC AAGGGGCTGA AGGATGCCCA GAAGGTACCC  CATTGTATGG GATCTGATCT GGGGCCTCGG TGCACATGCT TTACATGTGT 2000
2001 TTAGTCGAGG TTAAAAAACG TCTAGGCCCC CCGAACCACG GGGACGTGGT  TTTCCTTTGA AAAACACGAT AATACCATGG CGCCCATCAC GGCGTACGCC 2100
2101 CAGCAGACGA GAGGCCTCCT AGGGTGTATA ATCACCAGCC TGACTGGCCG  GGACAAAAAC CAAGTGGAGG GTGAGGTCCA GATCGTGTCA ACTGCTACCC 2200
2201 AAACCTTCCT GGCAACGTGC ATCAATGGGG TATGCTGGAC TGTCTACCAC  GGGGCCGGAA CGAGGACCAT CGCATCACCC AAGGGTCCTG TCATCCAGAT 2300
2301 GTATACCAAT GTGGACCAAG ACCTTGTGGG CTGGCCCGCT CCTCAAGGTT  CCCGCTCATT GACACCCTGT ACCTGCGGCT CCTCGGACCT TTACCTGGTC 2400
2401 ACGAGGCACG CCGATGTCAT TCCCGTGCGC CGGCGAGGTG ATAGCAGGGG  TAGCCTGCTT TCGCCCCGGC CCATTTCCTA CTTGAAAGGC TCCTCGGGGG 2500
2501 GTCCGCTGTT GTGCCCCGCG GGACACGCCG TGGGCCTATT CAGGGCCGCG  GTGTGCACCC GTGGAGTGGC TAAAGCGGTG GACTTTATCC CTGTGGAGAA 2600
2601 CCTAGGGACA ACCATGAGAT CCCCGGTGTT CACGGACAAC TCCTCTCCAC  CAGCAGTGCC CCAGAGCTTC CAGGTGGCCC ACCTGCATGC TCCCACCGGC 2700
2701 AGCGGTAAGA GCACCAAGGT CCCGGCTGCG TACGCAGCCC AGGGCTACAA  GGTGTTGGTG CTCAACCCCT CTGTTGCTGC AACGCTGGGC TTTGGTGCTT 2800
2801 ACATGTCCAA GGCCCATGGG GTTGATCCTA ATATCAGGAC CGGGGTGAGA  ACAATTACCA CTGGCAGCCC CATCACGTAC TCCACCTACG GCAAGTTCCT 2900
2901 TGCCGACGGC GGGTGCTCAG GAGGTGCTTA TGACATAATA ATTTGTGACG  AGTGCCACTC CACGGATGCC ACATCCATCT TGGGCATCGG CACTGTCCTT 3000
3001 GACCAAGCAG AGACTGCGGG GGCGAGACTG GTTGTGCTCG CCACTGCTAC  CCCTCCGGGC TCCGTCACTG TGTCCCATCC TAACATCGAG GAGGTTGCTC 3100
3101 TGTCCACCAC CGGAGAGATC CCCTTTTACG GCAAGGCTAT CCCCCTCGAG  GTGATCAAGG GGGGAAGACA TCTCATCTTC TGCCACTCAA AGAAGAAGTG 3200
3201 CGACGAGCTC GCCGCGAAGC TGGTCGCATT GGGCATCAAT GCCGTGGCCT  ACTACCGCGG TCTTGACGTG TCTGTCATCC CGACCAGCGG CGATGTTGTC 3300
3301 GTCGTGTCGA CCGATGCTCT CATGACTGGC TTTACCGGCG ACTTCGACTC  TGTGATAGAC TGCAACACGT GTGTCACTCA GACAGTCGAT TTCAGCCTTG 3400
3401 ACCCTACCTT TACCATTGAG ACAACCACGC TCCCCCAGGA TGCTGTCTCC  AGGACTCAAC GCCGGGGCAG GACTGGCAGG GGGAAGCCAG GCATCTATAG 3500
3501 ATTTGTGGCA CCGGGGGAGC GCCCCTCCGG CATGTTCGAC TCGTCCGTCC  TCTGTGAGTG CTATGACGCG GGCTGTGCTT GGTATGAGCT CACGCCCGCC 3600
3601 GAGACTACAG TTAGGCTACG AGCGTACATG AACACCCCGG GGCTTCCCGT  GTGCCAGGAC CATCTTGAAT TTTGGGAGGG CGTCTTTACG GGCCTCACTC 3700
3701 ATATAGATGC CCACTTTTTA TCCCAGACAA AGCAGAGTGG GGAGAACTTT  CCTTACCTGG TAGCGTACCA AGCCACCGTG TGCGCTAGGG CTCAAGCCCC 3800
3801 TCCCCCATCG TGGGACCAGA TGTGGAAGTG TTTGATCCGC CTTAAACCCA  CCCTCCATGG GCCAACACCC CTGCTATACA GACTGGGCGC TGTTCAGAAT 3900
3901 GAAGTCACCC TGACGCACCC AATCACCAAA TACATCATGA CATGCATGTC  GGCCGACCTG GAGGTCGTCA CGAGCACCTG GGTGCTCGTT GGCGGCGTCC 4000
4001 TGGCTGCTCT GGCCGCGTAT TGCCTGTCAA CAGGCTGCGT GGTCATAGTG  GGCAGGATCG TCTTGTCCGG GAAGCCGGCA ATTATACCTG ACAGGGAGGT 4100
```

Fig. 12A-2

```
4101 TCTCTACCAG GAGTTCGATG AGATGGAAGA GTGCTCTCAG CACTTACCGT ACATCGAGCA AGGGATGATG CTCGCTGAGC AGTTCAAGCA GAAGGCCCTC 4200
4201 GGCCTCCTGC AGACCGCGTC CCGCCATGCA GAGGTTATCA CCCCTGCTGT CCAGACCAAC TGGCAGAAAC TCGAGGTCTT TTGGGCGAAG CACATGTGGA 4300
4301 ATTTCATCAG TGGGATACAA TACTTGGCGG GCCTGTCAAC GCTGCCTGGT AACCCCGCCA TTGCTTCATT GATGGCTTTT ACAGCTGCCG TCACCAGCCC 4400
4401 ACTAACCACT GGCCAAACCC TCCTCTTCAA CATATTGGGG GGGTGGGTGG CTGCCCAGCT CGCCGCCCCC GGTGCCGCTA CTGCCTTTGT GGGTGCTGGC 4500
4501 CTAGCTGGCG CCGCCATCGG CAGCGTTGGA CTGGGGAAGG TCCTCGTGGA CATTCTTGCA GGGTATGGCG CGGGCGTGGC GGGAGCTCTT GTAGCATTCA 4600
4601 AGATCATGAG CGGTGAGGTC CCCTCCACGG AGGACCTGGT CAATCTGCTG CCCGCCATCC TCTCGCCTGG AGCCCTTGTA GTCGGTGTGG TCTGCGCAGC 4700
4701 AATACTGCGC CGGCACGTTG GCCCGGGCGA GGGGGCAGTG CAATGGATGA ACCGGCTAAT AGCCTTCGCC TCCCGGGGGA ACCATGTTTC CCCCACGCAC 4800
4801 TACGTGCCGG AGAGCGATGC AGCCGCCCGC GTCACTGCCA TACTCAGCAG CCTCACTGTA ACCCAGCTCC TGAGGCGACT GCATCAGTGG ATAAGCTCGG 4900
4901 AGTGTACCAC TCCATGCTCC GGTTCCTGGC TAAGGGACAT CTGGGACTGG ATATGCGAGG TGCTGAGCGA CTTTAAGACC TGGCTGAAAG CCAAGCTCAT 5000
5001 GCCACAACTG CCTGGGATTC CCTTTGTGTC CTGCCAGCGC GGGTATAGGG GGGTCTGGCG AGGAGACGGC ATTATGCACA CTCGCTGCCA CTGTGGAGCT 5100
5101 GAGATCACTG GACATGTCAA AAACGGGACG ATGAGGATCG TCGGTCCTAG GACCTGCAGG AACATGTGGA GTGGGACGTT CCCCATTAAC GCCTACACCA 5200
5201 CGGGCCCCTG TACTCCCCTT CCTGCGCCGA ACTATAAGTT CGCGCTGTGG AGGGTGTCTG CAGAGGAATA CGTGGAGATA AGGCGGGTGG GGGACTTCCA 5300
5301 CTACGTATCG GGTATGACTA CTGACAATCT TAAATGCCCG TGCCAGATCC CATCGCCCGA ATTTTTCACA GAATTGGACG GGGTGCGCCT ACACAGGTTT 5400
5401 GCGCCCCCTT GCAAGCCCTT GCTGCGGGAG GAGGTATCAT TCAGAGTAGG ACTCCACGAG TACCCGGTGG GGTCGCAATT ACCTTGCGAG CCCGAACCGG 5500
5501 ACGTAGCCGT GTTGACGTCC ATGCTCACTG ATCCCTCCCA TATAACAGCA GAGGCGGCCG GGAGAAGGTT GGCGAGAGGG TCACCCCCTT CTATGGCCAG 5600
5601 CTCCTCGGCT AGCCAGCTGT CCGCTCCATC TCTCAAGGCA ACTTGCACCG CCAACCATGA CTCCCCTGAC GCCGAGCTCA TAGAGGCTAA CCTCCTGTGG 5700
5701 AGGCAGGAGA TGGGCGGCAA CATCACCAGG GTTGAGTCAG AGAACAAAGT GGTGATTCTG GACTCCTTCG ATCCGCTTGT GGCAGAGGAG GATGAGCGGG 5800
5801 AGGTCTCCGT ACCTGCAGAA ATTCTGCGGA AGTCTCGGAG ATTCGCCCGG GCCCTGCCCG TCTGGGCGCG GCCGGACTAC AACCCCCCGC TAGTAGAGAC 5900
5901 GTGGAAAAAG CCTGACTACG AACCACCTGT GGTCCATGGC TGCCCGCTAC CACCTCCACG GTCCCCTCCT GTGCCTCCGC CTCGGAAAAA GCGTACGGTG 6000
6001 GTCCTCACCG AATCAACCCT ATCTACTGCC TTGGCCGAGC TTGCCACCAA AAGTTTTGGC AGCTCCTCAA CTTCCGGCAT TACGGGCGAC AATACGACAA 6100
6101 CATCCTCTGA GCCCGCCCCT TCTGGCTGCC CCCCCGACTC CGACGTTGAG TCCTATTCTT CCATGCCCCC CCTGGAGGGG GAGCCTGGGG ATCCGGATCT 6200
6201 CAGCGACGGG TCATGGTCGA CGGTCAGTAG TGGGGCCGAC ACGGAAGATG TCGTGTGCTG CTCAATGTCT TATTCCTGGA CAGGCGCACT CGTCACCCCG 6300
6301 TGCGCTGCGG AAGAACAAAA ACTGCCCATC AACGCACTGA GCAACTCGTT GCTACGCCAT CACAATCTGG TGTATTCCAC CACTTCACGC AGTGCTTGCC 6400
6401 AAAGGCAGAA GAAAGTCACA TTTGACAGAC TGCAAGTTCT GGACAGCCAT TACCAGGACG TGCTCAAGGA GGTCAAAGCA GCGGCGTCAA AAGTGAAGGC 6500
6501 TAACTTGCTA TCCGTAGAGG AAGCTTGCAG CCTGACGCCC CCACATTCAG CCAAATCCAA GTTTGGCTAT GGGGCAAAAG ACGTCCGTTG CCATGCCAGA 6600
6601 AAGGCCGTAG CCCACATCAA CTCCGTGTGG AAAGACCTTC TGGAAGACAG TGTAACACCA ATAGACACTA CCATCATGGC CAAGAACGAG GTTTTCTGCG 6700
6701 TTCAGCCTGA GAAGGGGGGT CGTAAGCCAG CTCGTCTCAT CGTGTTCCCC GACCTGGGCG TGCGCGTGTG CGAGAAGATG GCCCTGTACG ACGTGGTTAG 6800
6801 CAAGCTCCCC CTGGCCGTGA TGGGAAGCTC CTACGGATTC CAATACTCAC CAGGACAGCG GGTTGAATTC CTCGTGCAAG CGTGGAAGTC CAAGAAGACC 6900
6901 CCGATGGGGT TCTCGTATGA TACCCGCTGT TTTGACTCCA CAGTCACTGA GAGCGACATC CGTACGGAGG AGGCAATTTA CCAATGTTGT GACCTGGACC 7000
7001 CCCAAGCCCG CGTGGCCATC AAGTCCCTCA CTGAGAGGCT TTATGTTGGG GGCCCTCTTA CCAATTCAAG GGGGGAAAAC TGCGGCTACC GCAGGTGCCG 7100
7101 CGCGAGCGGC GTACTGACAA CTAGCTGTGG TAACACCCTC ACTTGCTACA TCAAGGCCCG GGCAGCCTGT CGAGCCGCAG GGCTCCAGGA CTGCACCATG 7200
7201 CTCGTGTGTG GCGACGACTT AGTCGTTATC TGTGAAAGTG CGGGGGTCCA GGAGGACGCG GCGAGCCTGA GAGCCTTCAC GGAGGCTATG ACCAGGTACT 7300
7301 CCGCCCCCCC CGGGGACCCC CCACAACCAG AATACGACTT GGAGCTTATA ACATCATGCT CCTCCAACGT GTCAGTCGCC CACGACGGCG CTGGAAAGAG 7400
7401 GGTCTACTAC CTTACCCGTG ACCCTACAAC CCCCCTCGCG AGAGCCGCGT GGGAGACAGC AAGACACACT CCAGTCAATT CCTGGCTAGG CAACATAATC 7500
7501 ATGTTTGCCC CCACACTGTG GGCGAGGATG ATACTGATGA CCCATTTCTT TAGCGTCCTC ATAGCCAGGG ATCAGCTTGA ACAGGCTCTT AACTGTGAGA 7600
7601 TCTACGGAGC CTGCTACTCC ATAGAACCAC TGGATCTACC TCCAATCATT CAAAGACTCC ATGGCCTCAG CGCATTTTCA CTCCACAGTT ACTCTCCAGG 7700
7701 TGAAATCAAT AGGGTGGCCG CATGCCTCAG AAAACTTGGG GTCCCGCCCT TGCGAGCTTG GAGACACCGG GCCCGGAGCG TCCGCGCTAG GCTTCTGTCC 7800
7801 AGAGGAGGCA GGGCTGCCAT ATGTGGCAAG TACCTCTTCA ACTGGGCAGT AAGAACAAAG CTCAAACTCA CTCCAATAGC GGCCGCTGGC CGGCTGGACT 7900
7901 TGTCCGGTTG GTTCACGGCT GGCTACAGCG GGGGAGACAT TTATCACAGC GTGTCTCATG CCCGGCCCCG CTGGTTCTGG TTTTGCCTAC TCCTGCTCGC 8000
8001 TGCAGGGGTA GGCATCTACC TCCTCCCCAA CCGATGAAGG TTGGGGTAAA CACTCCGGCC TCTTAAGCCA TTTCCTGTTT TTTTTTTTTT TTTTTTTTTT 8100
8101 TTTTTCTTTT TTTTTTTCTT TCCTTTCCTT CTTTTTTTCC TTTCTTTTTC CCTTCTTTAA TGGTGGCTCC ATCTTAGCCC TAGTCACGGC TAGCTGTGAA 8200
8201 AGGTCCGTGA GCCGCATGAC TGCAGAGAGT GCTGATACTG GCCTCTCTGC AGATCATGTG GGTCGGCATG GCATCTCCAC CTCCTCGCGG TCCGACCTGG 8300
```

Fig. 12A-3

```
8301 GCATCCGAAG GAGGACGCAC GTCCACTCGG ATGGCTAAGG GAGTCTAGAC TGGAATTCGT CGACGAGCTC CCTATAGTGA GTCGTATTAG AGGCCGACTT 8400
8401 GGCCAAATTC GTAATCATGG TCATAGCTGT TTCCTGTGTG AAATTGTTAT CCGCTCACAA TTCCACACAA CATACGAGCC GGAAGCATAA AGTGTAAAGC 8500
8501 CTGGGGTGCC TAATGAGTGA GCTAACTCAC ATTAATTGCG TTGCGCTCAC TGCCCGCTTT CCAGTCGGGA AACCTGTCGT GCCAGCTGCA TTAATGAATC 8600
8601 GGCCAACGCG CGGGGAGAGG CGGTTTGCGT ATTGGGCGCT CTTCCGCTTC CTCGCTCACT GACTCGCTGC GCTCGGTCGT TCGGCTGCGG CGAGCGGTAT 8700
8701 CAGCTCACTC AAAGGCGGTA ATACGGTTAT CCACAGAATC AGGGGATAAC GCAGGAAAGA ACATGTGAGC AAAAGGCCAG CAAAAGGCCA GGAACCGTAA 8800
8801 AAAGGCCGCG TTGCTGGCGT TTTTCCATAG GCTCCGCCCC CCTGACGAGC ATCACAAAAA TCGACGCTCA AGTCAGAGGT GGCGAAACCC GACAGGACTA 8900
8901 TAAAGATACC AGGCGTTTCC CCCTGGAAGC TCCCTCGTGC GCTCTCCTGT TCCGACCCTG CCGCTTACCG GATACCTGTC CGCCTTTCTC CCTTCGGGAA 9000
9001 GCGTGGCGCT TTCTCATAGC TCACGCTGTA GGTATCTCAG TTCGGTGTAG GTCGTTCGCT CCAAGCTGGG CTGTGTGCAC GAACCCCCCG TTCAGCCCGA 9100
9101 CCGCTGCGCC TTATCCGGTA ACTATCGTCT TGAGTCCAAC CCGGTAAGAC ACGACTTATC GCCACTGGCA GCAGCCACTG GTAACAGGAT TAGCAGAGCG 9200
9201 AGGTATGTAG GCGGTGCTAC AGAGTTCTTG AAGTGGTGGC CTAACTACGG CTACACTAGA AGGACAGTAT TTGGTATCTG CGCTCTGCTG AAGCCAGTTA 9300
9301 CCTTCGGAAA AAGAGTTGGT AGCTCTTGAT CCGGCAAACA AACCACCGCT GGTAGCGGTG GTTTTTTTGT TTGCAAGCAG CAGATTACGC GCAGAAAAAA 9400
9401 AGGATCTCAA GAAGATCCTT TGATCTTTTC TACGGGGTCT GACGCTCAGT GGAACGAAAA CTCACGTTAA GGGATTTTGG TCATGAGATT ATCAAAAAGG 9500
9501 ATCTTCACCT AGATCCTTTT AAATTAAAAA TGAAGTTTTA AATCAATCTA AAGTATATAT GAGTAAACTT GGTCTGACAG TTACCAATGC TTAATCAGTG 9600
9601 AGGCACCTAT CTCAGCGATC TGTCTATTTC GTTCATCCAT AGTTGCCTGA CTCCCCGTCG TGTAGATAAC TACGATACGG GAGGGCTTAC CATCTGGCCC 9700
9701 CAGTGCTGCA ATGATACCGC GAGACCCACG CTCACCGGCT CCAGATTTAT CAGCAATAAA CCAGCCAGCC GGAAGGGCCG AGCGCAGAAG TGGTCCTGCA 9800
9801 ACTTTATCCG CCTCCATCCA GTCTATTAAT TGTTGCCGGG AAGCTAGAGT AAGTAGTTCG CCAGTTAATA GTTTGCGCAA CGTTGTTGCC ATTGCTACAG 9900
9901 GCATCGTGGT GTCACGCTCG TCGTTTGGTA TGGCTTCATT CAGCTCCGGT TCCCAACGAT CAAGGCGAGT TACATGATCC CCCATGTTGT GCAAAAAAGC 10000
10001 GGTTAGCTCC TTCGGTCCTC CGATCGTTGT CAGAAGTAAG TTGGCCGCAG TGTTATCACT CATGGTTATG GCAGCACTGC ATAATTCTCT TACTGTCATG 10100
10101 CCATCCGTAA GATGCTTTTC TGTGACTGGT GAGTACTCAA CCAAGTCATT CTGAGAATAG TGTATGCGGC GACCGAGTTG CTCTTGCCCG GCGTCAATAC 10200
10201 GGGATAATAC CGCGCCACAT AGCAGAACTT TAAAAGTGCT CATCATTGGA AAACGTTCTT CGGGGCGAAA ACTCTCAAGG ATCTTACCGC TGTTGAGATC 10300
10301 CAGTTCGATG TAACCCACTC GTGCACCCAA CTGATCTTCA GCATCTTTTA CTTTCACCAG CGTTTCTGGG TGAGCAAAAA CAGGAAGGCA AAATGCCGCA 10400
10401 AAAAAGGGAA TAAGGGCGAC ACGGAAATGT TGAATACTCA TACTCTTCCT TTTTCAATAT TATTGAAGCA TTTATCAGGG TTATTGTCTC ATGAGCGGAT 10500
10501 ACATATTTGA ATGTATTTAG AAAAATAAAC AAATAGGGGT TCCGCGCACA TTTCCCCGAA AAGTGCCACC TGACGCGCCC TGTAGCGGCG CATTAAGCGC 10600
10601 GGCGGGTGTG GTGGTTACGC GCAGCGTGAC CGCTACACTT GCCAGCGCCC TAGCGCCCGC TCCTTTCGCT TTCTTCCCTT CCTTTCTCGC CACGTTCGCC 10700
10701 GGCTTTCCCC GTCAAGCTCT AAATCGGGGC ATCCCTTTAG GGTTCCGATT TAGTGCTTTA CGGCACCTCG ACCCCAAAAA ACTTGATTAG GGTGATGGTT 10800
10801 CACGTAGTGG GCCATCGCCC TGATAGACGG TTTTTCGCCC TTTGACGTTG GAGTCCACGT TCTTTAATAG TGGACTCTTG TTCCAAACTG GAACAACACT 10900
10901 CAACCCTATC TCGGTCTATT CTTTTGATTT ATAAGGGATT TTGCCGATTT CGGCCTATTG GTTAAAAAAT GAGCTGATTT AACAAAAATT TAACGCGAAT 11000
11001 TTTAACAAAA TATTAACAAA ATATTAACGT TTACAATTTC CCATTCGCCA TTCAGGCTGC GCAACTGTTG GGAAGGGCGA TCGGTGCGGG CCTCTTCGCT 11100
11101 ATTACGCCAG CTGGCGAAAG GGGGATGTGC TGCAAGGCGA TTAAGTTGGG TAACGCCAGG GTTTTCCCAG TCACGACGTT GTAAAACGAC GGCCAGTGCC 11200
11201 AAGCTGACTT GGTCAGCGGC CGCTAATACG ACTCACTATA
```

Fig. 12B-1

```
2077/1                                  2107/11                                 2137/21
ATG GCG CCC ATC ACG GCG TAC GCC CAG CAG ACG AGA GGC CTC CTA GGG TGT ATA ATC ACC AGC CTG ACT GGC CGG GAC AAA AAC CAA GTG
M   A   P   I   T   A   Y   A   Q   Q   T   R   G   L   L   G   C   I   I   T   S   L   T   G   R   D   K   N   Q   V
2167/31                                 2197/41                                 2227/51
GAG GGT GAG GTC CAG ATC GTG TCA ACT GCT ACC CAA ACC TTC CTG GCA ACG TGC ATC AAT GGG GTA TGC TGG ACT GTC TAC CAC GGG GCC
E   G   E   V   Q   I   V   S   T   A   T   Q   T   F   L   A   T   C   I   N   G   V   C   W   T   V   Y   H   G   A
2257/61                                 2287/71                                 2317/81
GGA ACG AGG ACC ATC GCA TCA CCC AAG GGT CCT GTC ATC CAG ATG TAT ACC AAT GTG GAC CAA GAC CTT GTG GGC TGG CCC GCT CCT CAA
G   T   R   T   I   A   S   P   K   G   P   V   I   Q   M   Y   T   N   V   D   Q   D   L   V   G   W   P   A   P   Q
2347/91                                 2377/101                                2407/111
GGT TCC CGC TCA TTG ACA CCC TGT ACC TGC GGC TCC TCG GAC CTT TAC CTG GTC ACG AGG CAC GCC GAT GTC ATT CCC GTG CGC CGG CGA
G   S   R   S   L   T   P   C   T   C   G   S   S   D   L   Y   L   V   T   R   H   A   D   V   I   P   V   R   R   R
2437/121                                2467/131                                2497/141
GGT GAT AGC AGG GGT AGC CTG CTT TCG CCC CGG CCC ATT TCC TAC TTG AAA GGC TCC TCG GGG GGT CCG CTG TTG TGC CCC GCG GGA CAC
G   D   S   R   G   S   L   L   S   P   R   P   I   S   Y   L   K   G   S   S   G   G   P   L   L   C   P   A   G   H
2527/151                                2557/161                                2587/171
GCC GTG GGC CTA TTC AGG GCC GCG GTG TGC ACC CGT GGA GTG GCT AAA GCG GTG GAC TTT ATC CCT GTG GAG AAC CTA GGG ACA ACC ATG
A   V   G   L   F   R   A   A   V   C   T   R   G   V   A   K   A   V   D   F   I   P   V   E   N   L   G   T   T   M
2617/181                                2647/191                                2677/201
AGA TCC CCG GTG TTC ACG GAC AAC TCC TCT CCA CCA GCA GTG CCC CAG AGC TTC CAG GTG GCC CAC CTG CAT GCT CCC ACC GGC AGC GGT
R   S   P   V   F   T   D   N   S   S   P   P   A   V   P   Q   S   F   Q   V   A   H   L   H   A   P   T   G   S   G
2707/211                                2737/221                                2767/231
AAG AGC ACC AAG GTC CCG GCT GCG TAC GCA GCC CAG GGC TAC AAG GTG TTG GTG CTC AAC CCC TCT GTT GCT GCA ACG CTG GGC TTT GGT
K   S   T   K   V   P   A   A   Y   A   A   Q   G   Y   K   V   L   V   L   N   P   S   V   A   A   T   L   G   F   G
2797/241                                2827/251                                2857/261
GCT TAC ATG TCC AAG GCC CAT GGG GTT GAT CCT AAT ATC AGG ACC GGG GTG AGA ACA ATT ACC ACT GGC AGC CCC ATC ACG TAC TCC ACC
A   Y   M   S   K   A   H   G   V   D   P   N   I   R   T   G   V   R   T   I   T   T   G   S   P   I   T   Y   S   T
2887/271                                2917/281                                2947/291
TAC GGC AAG TTC CTT GCC GAC GGC GGG TGC TCA GGA GGT GCT TAT GAC ATA ATA ATT TGT GAC GAG TGC CAC TCC ACG GAT GCC ACA TCC
Y   G   K   F   L   A   D   G   G   C   S   G   G   A   Y   D   I   I   I   C   D   E   C   H   S   T   D   A   T   S
2977/301                                3007/311                                3037/321
ATC TTG GGC ATC GGC ACT GTC CTT GAC CAA GCA GAG ACT GCG GGG GCG AGA CTG GTT GTG CTC GCC ACT GCT ACC CCT CCG GGC TCC GTC
I   L   G   I   G   T   V   L   D   Q   A   E   T   A   G   A   R   L   V   V   L   A   T   A   T   P   P   G   S   V
3067/331                                3097/341                                3127/351
ACT GTG TCC CAT CCT AAC ATC GAG GAG GTT GCT CTG TCC ACC ACC GGA GAG ATC CCC TTT TAC GGC AAG GCT ATC CCC CTC GAG GTG ATC
T   V   S   H   P   N   I   E   E   V   A   L   S   T   T   G   E   I   P   F   Y   G   K   A   I   P   L   E   V   I
3157/361                                3187/371                                3217/381
AAG GGG GGA AGA CAT CTC ATC TTC TGC CAC TCA AAG AAG AAG TGC GAC GAG CTC GCC GCG AAG CTG GTC GCA TTG GGC ATC AAT GCC GTG
K   G   G   R   H   L   I   F   C   H   S   K   K   K   C   D   E   L   A   A   K   L   V   A   L   G   I   N   A   V
3247/391                                3277/401                                3307/411
GCC TAC TAC CGC GGT CTT GAC GTG TCT GTC ATC CCG ACC AGC GGC GAT GTT GTC GTC GTG TCG ACC GAT GCT CTC ATG ACT GGC TTT ACC
A   Y   Y   R   G   L   D   V   S   V   I   P   T   S   G   D   V   V   V   V   S   T   D   A   L   M   T   G   F   T
3337/421                                3367/431                                3397/441
GGC GAC TTC GAC TCT GTG ATA GAC TGC AAC ACG TGT GTC ACT CAG ACA GTC GAT TTC AGC CTT GAC CCT ACC TTT ACC ATT GAG ACA ACC
G   D   F   D   S   V   I   D   C   N   T   C   V   T   Q   T   V   D   F   S   L   D   P   T   F   T   I   E   T   T
```

Fig. 12B-2

```
3427/451                                3457/461                                3487/471
ACG CTC CCC CAG GAT GCT GTC TCC AGG ACT CAA CGC CGG GGC AGG ACT GGC AGG GGG AAG CCA GGC ATC TAT AGA TTT GTG GCA CCG GGG
T   L   P   Q   D   A   V   S   R   T   Q   R   R   G   R   T   G   R   G   K   P   G   I   Y   R   F   V   A   P   G
3517/481                                3547/491                                3577/501
GAG CGC CCC TCC GGC ATG TTC GAC TCG TCC GTC CTC TGT GAG TGC TAT GAC GCG GGC TGT GCT TGG TAT GAG CTC ACG CCC GCC GAG ACT
E   R   P   S   G   M   F   D   S   S   V   L   C   E   C   Y   D   A   G   C   A   W   Y   E   L   T   P   A   E   T
3607/511                                3637/521                                3667/531
ACA GTT AGG CTA CGA GCG TAC ATG AAC ACC CCG GGG CTT CCC GTG TGC CAG GAC CAT CTT GAA TTT TGG GAG GGC GTC TTT ACG GGC CTC
T   V   R   L   R   A   Y   M   N   T   P   G   L   P   V   C   Q   D   H   L   E   F   W   E   G   V   F   T   G   L
3697/541                                3727/551                                3757/561
ACT CAT ATA GAT GCC CAC TTT TTA TCC CAG ACA AAG CAG AGT GGG GAG AAC TTT CCT TAC CTG GTA GCG TAC CAA GCC ACC GTG TGC GCT
T   H   I   D   A   H   F   L   S   Q   T   K   Q   S   G   E   N   F   P   Y   L   V   A   Y   Q   A   T   V   C   A
3787/571                                3817/581                                3847/591
AGG GCT CAA GCC CCT CCC CCA TCG TGG GAC CAG ATG TGG AAG TGT TTG ATC CGC CTT AAA CCC ACC CTC CAT GGG CCA ACA CCC CTG CTA
R   A   Q   A   P   P   P   S   W   D   Q   M   W   K   C   L   I   R   L   K   P   T   L   H   G   P   T   P   L   L
3877/601                                3907/611                                3937/621
TAC AGA CTG GGC GCT GTT CAG AAT GAA GTC ACC CTG ACG CAC CCA ATC ACC AAA TAC ATC ATG ACA TGC ATG TCG GCC GAC CTG GAG GTC
Y   R   L   G   A   V   Q   N   E   V   T   L   T   H   P   I   T   K   Y   I   M   T   C   M   S   A   D   L   E   V
3967/631                                3997/641                                4027/651
GTC ACG AGC ACC TGG GTG CTC GTT GGC GGC GTC CTG GCT GCT CTG GCC GCG TAT TGC CTG TCA ACA GGC TGC GTG GTC ATA GTG GGC AGG
V   T   S   T   W   V   L   V   G   G   V   L   A   A   L   A   A   Y   C   L   S   T   G   C   V   V   I   V   G   R
4057/661                                4087/671                                4117/681
ATC GTC TTG TCC GGG AAG CCG GCA ATT ATA CCT GAC AGG GAG GTT CTC TAC CAG GAG TTC GAT GAG ATG GAA GAG TGC TCT CAG CAC TTA
I   V   L   S   G   K   P   A   I   I   P   D   R   E   V   L   Y   Q   E   F   D   E   M   E   E   C   S   Q   H   L
4147/691                                4177/701                                4207/711
CCG TAC ATC GAG CAA GGG ATG ATG CTC GCT GAG CAG TTC AAG CAG AAG GCC CTC GGC CTC CTG CAG ACC GCG TCC CGC CAT GCA GAG GTT
P   Y   I   E   Q   G   M   M   L   A   E   Q   F   K   Q   K   A   L   G   L   L   Q   T   A   S   R   H   A   E   V
4237/721                                4267/731                                4297/741
ATC ACC CCT GCT GTC CAG ACC AAC TGG CAG AAA CTC GAG GTC TTT TGG GCG AAG CAC ATG TGG AAT TTC ATC AGT GGG ATA CAA TAC TTG
I   T   P   A   V   Q   T   N   W   Q   K   L   E   V   F   W   A   K   H   M   W   N   F   I   S   G   I   Q   Y   L
4327/751                                4357/761                                4387/771
GCG GGC CTG TCA ACG CTG CCT GGT AAC CCC GCC ATT GCT TCA TTG ATG GCT TTT ACA GCT GCC GTC ACC AGC CCA CTA ACC ACT GGC CAA
A   G   L   S   T   L   P   G   N   P   A   I   A   S   L   M   A   F   T   A   A   V   T   S   P   L   T   T   G   Q
4417/781                                4447/791                                4477/801
ACC CTC CTC TTC AAC ATA TTG GGG GGG TGG GTG GCT GCC CAG CTC GCC GCC CCC GGT GCC GCT ACT GCC TTT GTG GGT GCT GGC CTA GCT
T   L   L   F   N   I   L   G   G   W   V   A   A   Q   L   A   A   P   G   A   A   T   A   F   V   G   A   G   L   A
4507/811                                4537/821                                4567/831
GGC GCC GCC ATC GGC AGC GTT GGA CTG GGG AAG GTC CTC GTG GAC ATT CTT GCA GGG TAT GGC GCG GGC GTG GCG GGA GCT CTT GTA GCA
G   A   A   I   G   S   V   G   L   G   K   V   L   V   D   I   L   A   G   Y   G   A   G   V   A   G   A   L   V   A
4597/841                                4627/851                                4657/861
TTC AAG ATC ATG AGC GGT GAG GTC CCC TCC ACG GAG GAC CTG GTC AAT CTG CTG CCC GCC ATC CTC TCG CCT GGA GCC CTT GTA GTC GGT
F   K   I   M   S   G   E   V   P   S   T   E   D   L   V   N   L   L   P   A   I   L   S   P   G   A   L   V   V   G
4687/871                                4717/881                                4747/891
GTG GTC TGC GCA GCA ATA CTG CGC CGG CAC GTT GGC CCG GGC GAG GGG GCA GTG CAA TGG ATG AAC CGG CTA ATA GCC TTC GCC TCC CGG
V   V   C   A   A   I   L   R   R   H   V   G   P   G   E   G   A   V   Q   W   M   N   R   L   I   A   F   A   S   R
```

Fig. 12B-3

```
4777/901                                4807/911                                4837/921
GGG AAC CAT GTT TCC CCC ACG CAC TAC GTG CCG GAG AGC GAT GCA GCC GCC CGC GTC ACT GCC ATA CTC AGC AGC CTC ACT GTA ACC CAG
G   N   H   V   S   P   T   H   Y   V   P   E   S   D   A   A   A   R   V   T   A   I   L   S   S   L   T   V   T   Q
4867/931                                4897/941                                4927/951
CTC CTG AGG CGA CTG CAT CAG TGG ATA AGC TCG GAG TGT ACC ACT CCA TGC TCC GGT TCC TGG CTA AGG GAC ATC TGG GAC TGG ATA TGC
L   L   R   R   L   H   Q   W   I   S   S   E   C   T   T   P   C   S   G   S   W   L   R   D   I   W   D   W   I   C
4957/961                                4987/971                                5017/981
GAG GTG CTG AGC GAC TTT AAG ACC TGG CTG AAA GCC AAG CTC ATG CCA CAA CTG CCT GGG ATT CCC TTT GTG TCC TGC CAG CGC GGG TAT
E   V   L   S   D   F   K   T   W   L   K   A   K   L   M   P   Q   L   P   G   I   P   F   V   S   C   Q   R   G   Y
5047/991                                5077/1001                               5107/1011
AGG GGG GTC TGG CGA GGA GAC GGC ATT ATG CAC ACT CGC TGC CAC TGT GGA GCT GAG ATC ACT GGA CAT GTC AAA AAC GGG ACG ATG AGG
R   G   V   W   R   G   D   G   I   M   H   T   R   C   H   C   G   A   E   I   T   G   H   V   K   N   G   T   M   R
5137/1021                               5167/1031                               5197/1041
ATC GTC GGT CCT AGG ACC TGC AGG AAC ATG TGG AGT GGG ACG TTC CCC ATT AAC GCC TAC ACC ACG GGC CCC TGT ACT CCC CTT CCT GCG
I   V   G   P   R   T   C   R   N   M   W   S   G   T   F   P   I   N   A   Y   T   T   G   P   C   T   P   L   P   A
5227/1051                               5257/1061                               5287/1071
CCG AAC TAT AAG TTC GCG CTG TGG AGG GTG TCT GCA GAG GAA TAC GTG GAG ATA AGG CGG GTG GGG GAC TTC CAC TAC GTA TCG GGT ATG
P   N   Y   K   F   A   L   W   R   V   S   A   E   E   Y   V   E   I   R   R   V   G   D   F   H   Y   V   S   G   M
5317/1081                               5347/1091                               5377/1101
ACT ACT GAC AAT CTT AAA TGC CCG TGC CAG ATC CCA TCG CCC GAA TTT TTC ACA GAA TTG GAC GGG GTG CGC CTA CAC AGG TTT GCG CCC
T   T   D   N   L   K   C   P   C   Q   I   P   S   P   E   F   F   T   E   L   D   G   V   R   L   H   R   F   A   P
5407/1111                               5437/1121                               5467/1131
CCT TGC AAG CCC TTG CTG CGG GAG GAG GTA TCA TTC AGA GTA GGA CTC CAC GAG TAC CCG GTG GGG TCG CAA TTA CCT TGC GAG CCC GAA
P   C   K   P   L   L   R   E   E   V   S   F   R   V   G   L   H   E   Y   P   V   G   S   Q   L   P   C   E   P   E
5497/1141                               5527/1151                               5557/1161
CCG GAC GTA GCC GTG TTG ACG TCC ATG CTC ACT GAT CCC TCC CAT ATA ACA GCA GAG GCG GCC GGG AGA AGG TTG GCG AGA GGG TCA CCC
P   D   V   A   V   L   T   S   M   L   T   D   P   S   H   I   T   A   E   A   A   G   R   R   L   A   R   G   S   P
5587/1171                               5617/1181                               5647/1191
CCT TCT ATG GCC AGC TCC TCG GCT AGC CAG CTG TCC GCT CCA TCT CTC AAG GCA ACT TGC ACC GCC AAC CAT GAC TCC CCT GAC GCC GAG
P   S   M   A   S   S   S   A   S   Q   L   S   A   P   S   L   K   A   T   C   T   A   N   H   D   S   P   D   A   E
5677/1201                               5707/1211                               5737/1221
CTC ATA GAG GCT AAC CTC CTG TGG AGG CAG GAG ATG GGC GGC AAC ATC ACC AGG GTT GAG TCA GAG AAC AAA GTG GTG ATT CTG GAC TCC
L   I   E   A   N   L   L   W   R   Q   E   M   G   G   N   I   T   R   V   E   S   E   N   K   V   V   I   L   D   S
5767/1231                               5797/1241                               5827/1251
TTC GAT CCG CTT GTG GCA GAG GAG GAT GAG CGG GAG GTC TCC GTA CCT GCA GAA ATT CTG CGG AAG TCT CGG AGA TTC GCC CGG GCC CTG
F   D   P   L   V   A   E   E   D   E   R   E   V   S   V   P   A   E   I   L   R   K   S   R   R   F   A   R   A   L
5857/1261                               5887/1271                               5917/1281
CCC GTC TGG GCG CGG CCG GAC TAC AAC CCC CCG CTA GTA GAG ACG TGG AAA AAG CCT GAC TAC GAA CCA CCT GTG GTC CAT GGC TGC CCG
P   V   W   A   R   P   D   Y   N   P   P   L   V   E   T   W   K   K   P   D   Y   E   P   P   V   V   H   G   C   P
5947/1291                               5977/1301                               6007/1311
CTA CCA CCT CCA CGG TCC CCT CCT GTG CCT CCG CCT CGG AAA AAG CGT ACG GTG GTC CTC ACC GAA TCA ACC CTA TCT ACT GCC TTG GCC
L   P   P   P   R   S   P   P   V   P   P   P   R   K   K   R   T   V   V   L   T   E   S   T   L   S   T   A   L   A
6037/1321                               6067/1331                               6097/1341
GAG CTT GCC ACC AAA AGT TTT GGC AGC TCC TCA ACT TCC GGC ATT ACG GGC GAC AAT ACG ACA ACA TCC TCT GAG CCC GCC CCT TCT GGC
E   L   A   T   K   S   F   G   S   S   S   T   S   G   I   T   G   D   N   T   T   T   S   S   E   P   A   P   S   G
```

Fig. 12B-4

```
6127/1351                                         6157/1361                                         6187/1371
TGC CCC CCC GAC TCC GAC GTT GAG TCC TAT TCT TCC ATG CCC CCC CTG GAG GGG GAG CCT GGG GAT CCG GAT CTC AGC GAC GGG TCA TGG
C   P   P   D   S   D   V   E   S   Y   S   S   M   P   P   L   E   G   E   P   G   D   P   D   L   S   D   G   S   W
6217/1381                                         6247/1391                                         6277/1401
TCG ACG GTC AGT AGT GGG GCC GAC ACG GAA GAT GTC GTG TGC TGC TCA ATG TCT TAT TCC TGG ACA GGC GCA CTC GTC ACC CCG TGC GCT
S   T   V   S   S   G   A   D   T   E   D   V   V   C   C   S   M   S   Y   S   W   T   G   A   L   V   T   P   C   A
6307/1411                                         6337/1421                                         6367/1431
GCG GAA GAA CAA AAA CTG CCC ATC AAC GCA CTG AGC AAC TCG TTG CTA CGC CAT CAC AAT CTG GTG TAT TCC ACC ACT TCA CGC AGT GCT
A   E   E   Q   K   L   P   I   N   A   L   S   N   S   L   L   R   H   H   N   L   V   Y   S   T   T   S   R   S   A
6397/1441                                         6427/1451                                         6457/1461
TGC CAA AGG CAG AAG AAA GTC ACA TTT GAC AGA CTG CAA GTT CTG GAC AGC CAT TAC CAG GAC GTG CTC AAG GAG GTC AAA GCA GCG GCG
C   Q   R   Q   K   K   V   T   F   D   R   L   Q   V   L   D   S   H   Y   Q   D   V   L   K   E   V   K   A   A   A
6487/1471                                         6517/1481                                         6547/1491
TCA AAA GTG AAG GCT AAC TTG CTA TCC GTA GAG GAA GCT TGC AGC CTG ACG CCC CCA CAT TCA GCC AAA TCC AAG TTT GGC TAT GGG GCA
S   K   V   K   A   N   L   L   S   V   E   E   A   C   S   L   T   P   P   H   S   A   K   S   K   F   G   Y   G   A
6577/1501                                         6607/1511                                         6637/1521
AAA GAC GTC CGT TGC CAT GCC AGA AAG GCC GTA GCC CAC ATC AAC TCC GTG TGG AAA GAC CTT CTG GAA GAC AGT GTA ACA CCA ATA GAC
K   D   V   R   C   H   A   R   K   A   V   A   H   I   N   S   V   W   K   D   L   L   E   D   S   V   T   P   I   D
6667/1531                                         6697/1541                                         6727/1551
ACT ACC ATC ATG GCC AAG AAC GAG GTT TTC TGC GTT CAG CCT GAG AAG GGG GGT CGT AAG CCA GCT CGT CTC ATC GTG TTC CCC GAC CTG
T   T   I   M   A   K   N   E   V   F   C   V   Q   P   E   K   G   G   R   K   P   A   R   L   I   V   F   P   D   L
6757/1561                                         6787/1571                                         6817/1581
GGC GTG CGC GTG TGC GAG AAG ATG GCC CTG TAC GAC GTG GTT AGC AAG CTC CCC CTG GCC GTG ATG GGA AGC TCC TAC GGA TTC CAA TAC
G   V   R   V   C   E   K   M   A   L   Y   D   V   V   S   K   L   P   L   A   V   M   G   S   S   Y   G   F   Q   Y
6847/1591                                         6877/1601                                         6907/1611
TCA CCA GGA CAG CGG GTT GAA TTC CTC GTG CAA GCG TGG AAG TCC AAG AAG ACC CCG ATG GGG TTC TCG TAT GAT ACC CGC TGT TTT GAC
S   P   G   Q   R   V   E   F   L   V   Q   A   W   K   S   K   K   T   P   M   G   F   S   Y   D   T   R   C   F   D
6937/1621                                         6967/1631                                         6997/1641
TCC ACA GTC ACT GAG AGC GAC ATC CGT ACG GAG GAG GCA ATT TAC CAA TGT TGT GAC CTG GAC CCC CAA GCC CGC GTG GCC ATC AAG TCC
S   T   V   T   E   S   D   I   R   T   E   E   A   I   Y   Q   C   C   D   L   D   P   Q   A   R   V   A   I   K   S
7027/1651                                         7057/1661                                         7087/1671
CTC ACT GAG AGG CTT TAT GTT GGG GGC CCT CTT ACC AAT TCA AGG GGG GAA AAC TGC GGC TAC CGC AGG TGC CGC GCG AGC GGC GTA CTG
L   T   E   R   L   Y   V   G   G   P   L   T   N   S   R   G   E   N   C   G   Y   R   R   C   R   A   S   G   V   L
7117/1681                                         7147/1691                                         7177/1701
ACA ACT AGC TGT GGT AAC ACC CTC ACT TGC TAC ATC AAG GCC CGG GCA GCC TGT CGA GCC GCA GGG CTC CAG GAC TGC ACC ATG CTC GTG
T   T   S   C   G   N   T   L   T   C   Y   I   K   A   R   A   A   C   R   A   A   G   L   Q   D   C   T   M   L   V
7207/1711                                         7237/1721                                         7267/1731
TGT GGC GAC GAC TTA GTC GTT ATC TGT GAA AGT GCG GGG GTC CAG GAG GAC GCG GCG AGC CTG AGA GCC TTC ACG GAG GCT ATG ACC AGG
C   G   D   D   L   V   V   I   C   E   S   A   G   V   Q   E   D   A   A   S   L   R   A   F   T   E   A   M   T   R
7297/1741                                         7327/1751                                         7357/1761
TAC TCC GCC CCC CCC GGG GAC CCC CCA CAA CCA GAA TAC GAC TTG GAG CTT ATA ACA TCA TGC TCC TCC AAC GTG TCA GTC GCC CAC GAC
Y   S   A   P   P   G   D   P   P   Q   P   E   Y   D   L   E   L   I   T   S   C   S   S   N   V   S   V   A   H   D
7387/1771                                         7417/1781                                         7447/1791
GGC GCT GGA AAG AGG GTC TAC TAC CTT ACC CGT GAC CCT ACA ACC CCC CTC GCG AGA GCC GCG TGG GAG ACA GCA AGA CAC ACT CCA GTC
G   A   G   K   R   V   Y   Y   L   T   R   D   P   T   T   P   L   A   R   A   A   W   E   T   A   R   H   T   P   V
```

Fig. 12B-5

```
7477/1801                                         7507/1811                                         7537/1821
AAT TCC TGG CTA GGC AAC ATA ATC ATG TTT GCC CCC ACA CTG TGG GCG AGG ATG ATA CTG ATG ACC CAT TTC TTT AGC GTC CTC ATA GCC
N   S   W   L   G   N   I   I   M   F   A   P   T   L   W   A   R   M   I   L   M   T   H   F   F   S   V   L   I   A
7567/1831                                         7597/1841                                         7627/1851
AGG GAT CAG CTT GAA CAG GCT CTT AAC TGT GAG ATC TAC GGA GCC TGC TAC TCC ATA GAA CCA CTG GAT CTA CCT CCA ATC ATT CAA AGA
R   D   Q   L   E   Q   A   L   N   C   E   I   Y   G   A   C   Y   S   I   E   P   L   D   L   P   P   I   I   Q   R
7657/1861                                         7687/1871                                         7717/1881
CTC CAT GGC CTC AGC GCA TTT TCA CTC CAC AGT TAC TCT CCA GGT GAA ATC AAT AGG GTG GCC GCA TGC CTC AGA AAA CTT GGG GTC CCG
L   H   G   L   S   A   F   S   L   H   S   Y   S   P   G   E   I   N   R   V   A   A   C   L   R   K   L   G   V   P
7747/1891                                         7777/1901                                         7807/1911
CCC TTG CGA GCT TGG AGA CAC CGG GCC CGG AGC GTC CGC GCT AGG CTT CTG TCC AGA GGA GGC AGG GCT GCC ATA TGT GGC AAG TAC CTC
P   L   R   A   W   R   H   R   A   R   S   V   R   A   R   L   L   S   R   G   G   R   A   A   I   C   G   K   Y   L
7837/1921                                         7867/1931                                         7897/1941
TTC AAC TGG GCA GTA AGA ACA AAG CTC AAA CTC ACT CCA ATA GCG GCC GCT GGC CGG CTG GAC TTG TCC GGT TGG TTC ACG GCT GGC TAC
F   N   W   A   V   R   T   K   L   K   L   T   P   I   A   A   A   G   R   L   D   L   S   G   W   F   T   A   G   Y
7927/1951                                         7957/1961                                         7987/1971
AGC GGG GGA GAC ATT TAT CAC AGC GTG TCT CAT GCC CGG CCC CGC TGG TTC TGG TTT TGC CTA CTC CTG CTC GCT GCA GGG GTA GGC ATC
S   G   G   D   I   Y   H   S   V   S   H   A   R   P   R   W   F   W   F   C   L   L   L   L   A   A   G   V   G   I
8017/1981                                         8047/1991                                         8077/2001
TAC CTC CTC CCC AAC CGA TGA
Y   L   L   P   N   R   *
```

Fig. 13A

```
   1 gccagccccc tgatgggggc gacactccac catgaatcac tcccctgtga ggaactactg
  61 tcttcacgca gaaagcgtct agccatggcg ttagtatgag tgtcgtgcag cctccaggac
 121 ccccccctccc gggagagcca tagtggtctg cggaaccggt gagtacaccg gaattgccag
 181 gacgaccggg tcctttcttg gataaacccg ctcaatgcct ggagatttgg gcgtgccccc
 241 gcaagactgc tagccgagta gtgttgggtc gcgaaaggcc ttgtggtact gcctgatagg
 301 gtgcttgcga gtgccccggg aggtctcgta gaccgtgcac catgagcacg aatcctaaac
 361 ctcaaagaaa aaccaaacgt aacaccaacc gtcgcccaca ggacgtcaag ttcccgggtg
 421 gcggtcagat cgttggtgga gtttacttgt tgccgcgcag gggccctaga ttgggtgtgc
 481 gcgcgacgag gaagacttcc gagcggtcgc aacctcgagg tagacgtcag cctatcccca
 541 aggcacgtcg gcccgagggc aggacctggg ctcagcccgg gtacccttgg cccctctatg
 601 gcaatgaggg ttgcgggtgg gcgggatggc tcctgtctcc ccgtggctct cggcctagct
 661 ggggccccac agacccccgg cgtaggtcgc gcaatttggg taaggtcatc gatacccttta
 721 cgtgcggctt cgccgacctc atggggtaca taccgctcgt cggcgcccct cttggaggcg
 781 ctgccagggc cctggcgcat ggcgtccggg ttctggaaga cggcgtgaac tatgcaacag
 841 ggaaccttcc tggttgctct ttctctatct tccttctggc cctgctctct tgcctgactg
 901 tgcccgcttc agcctaccaa gtgcgcaatt cctcggggct ttaccatgtc accaatgatt
 961 gccctaactc gagtattgtg tacgaggcgg ccgatgccat cctgcacact ccggggtgtg
1021 tcccttgcgt tcgcgagggt aacgcctcga ggtgttgggt ggcggtgacc cccacggtgg
1081 ccaccaggga cggcaaactc cccacaacgc agcttcgacg tcatatcgat ctgcttgtcg
1141 ggagcgccac cctctgctcg gccctctacg tgggggacct gtgcgggtct gtctttcttg
1201 ttggtcaact gtttaccttc tctcccaggc gccactggac gacgcaagac tgcaattgtt
1261 ctatctatcc cggccatata acgggtcatc gcatggcatg ggatatgatg atgaactggt
1321 cccctacggc agcgttggtg gtagctcagc tgctccggat cccacaagcc atcatggaca
1381 tgatcgctgg tgctcactgg ggagtcctgg cgggcatagc gtatttctcc atggtgggga
1441 actgggcgaa ggtcctggta gtgctgctgc tatttgccgg cgtcgacgcg gaaacccacg
1501 tcaccggggg aaatgccggc cgcaccacgg ctgggcttgt tggtctcctt acaccaggcg
1561 ccaagcagaa catccaactg atcaacacca acggcagttg gcacatcaat agcacggcct
1621 tgaattgcaa tgaaagcctt aacaccggct ggttagcagg gctcttctat caacacaaat
1681 tcaactcttc aggctgtcct gagaggttgg ccagctgccg acgccttacc gattttgccc
1741 agggctgggg tcctatcagt tatgccaacg gaagcggcct cgacgaacgc ccctactgct
1801 ggcactaccc tccaagacct tgtggcattg tgcccgcaaa gagcgtgtgt ggcccggtat
1861 attgcttcac tcccagcccc gtggtggtgg gaacgaccga caggtcgggc gcgcctacct
1921 acagctgggg tgcaaatgat acggatgtct tcgtccttaa caacaccagg ccaccgctgg
1981 gcaattggtt cggttgtacc tggatgaact caactggatt caccaaagtg tgcggagcgc
2041 cccccttgtgt catcggaggg gtgggcaaca acacccttgct ctgccccact gattgcttcc
2101 gcaaacatcc ggaagccaca tactctcggt gcggctccgg tccctggatt acacccaggt
2161 gcatggtcga ctacccgtat aggctttggc actatccttg taccatcaat tacaccatat
2221 tcaaagtcag gatgtacgtg ggaggggtcg agcacaggct ggaagcggcc tgcaactgga
2281 cgcggggcga acgctgtgat ctggaagaca gggacaggtc cgagctcagc ccgttgctgc
2341 tgtccaccac acagtggcag gtccttccgt gttctttcac gaccctgcca gccttgtcca
2401 ccggcctcat ccacctccac cagaacattg tggacgtgca gtacttgtac ggggtagggt
2461 caagcatcgc gtcctgggcc attaagtggg agtacgtcgt tctcctgttc cttctgcttg
2521 cagacgcgcg cgtctgctcc tgcttgtgga tgatgttact catatcccaa gcggaggcgg
2581 ctttggagaa cctcgtaata ctcaatgcag catccctggc cgggacgcac ggtcttgtgt
2641 ccttcctcgt gttcttctgc tttgcgtggt atctgaaggg taggtgggtg cccggagcgg
2701 tctacgccct ctacgggatg tggcctctcc tcctgctcct gctggcgttg cctcagcggg
2761 catacgcact ggacacggag gtggccgcgt cgtgtggcgg cgttgttctt gtcgggttaa
2821 tggcgctgac tctgtcgcca tattacaagc gctatatcag ctggtgcatg tggtggcttc
2881 agtattttct gaccagagta gaagcgcaac tgcacgtgtg ggttcccccc ctcaacgtcc
2941 gggggggcg cgatgccgtc atcttactca tgtgtgtagt acacccgacc ctggtatttg
3001 acatcaccaa actactcctg gccatcttcg gacccctttg gattcttcaa gccagtttgc
3061 ttaaagtccc ctacttcgtg cgcgttcaag gccttctccg gatctgcgcg ctagcgcgga
```

Fig. 13B


```
3121 agatagccgg aggtcattac gtgcaaatgg ccatcatcaa gttaggggcg cttactggca
3181 cctatgtgta taaccatctc acccctcttc gagactgggc gcacaacggc ctgcgagatc
3241 tggccgtggc tgtggaacca gtcgtcttct cccgaatgga gaccaagctc atcacgtggg
3301 gggcagatac cgccgcgtgc ggtgacatca tcaacggctt gcccgtctct gcccgtaggg
3361 gccaggagat actgcttggg ccagccgacg gaatggtctc caaggggtgg aggttgctgg
3421 cgcccatcac ggcgtacgcc cagcagacga gaggcctcct agggtgtata atcaccagcc
3481 tgactggccg ggacaaaaac caagtggagg gtgaggtcca gatcgtgtca actgctaccc
3541 aaaccttcct ggcaacgtgc atcaatgggg tatgctggac tgtctaccac ggggccggaa
3601 cgaggaccat cgcatcaccc aagggtcctg tcatccagat gtataccaat gtggaccaag
3661 accttgtggg ctggcccgct cctcaaggtt cccgctcatt gacaccctgt acctgcggct
3721 cctcggacct ttacctggtc acgaggcacg ccgatgtcat tcccgtgcgc cggcgaggtg
3781 atagcagggg tagcctgctt tcgccccggc ccatttccta cttgaaaggc tcctcggggg
3841 gtccgctgtt gtgccccgcg ggacacgccg tgggcctatt cagggccgcg gtgtgcaccc
3901 gtggagtggc taaagcggtg gactttatcc ctgtggagaa cctagggaca accatgagat
3961 ccccggtgtt cacggacaac tcctctccac cagcagtgcc ccagagcttc caggtggccc
4021 acctgcatgc tcccaccggc agcggtaaga gcaccaaggt cccggctgcg tacgcagccc
4081 agggctacaa ggtgttggtg ctcaacccct ctgttgctgc aacgctgggc tttggtgctt
4141 acatgtccaa ggcccatggg gttgatccta atatcaggac cggggtgaga acaattacca
4201 ctggcagccc catcacgtac tccacctacg gcaagttcct tgccgacggc gggtgctcag
4261 gaggtgctta tgacataata atttgtgacg agtgccactc cacggatgcc acatccatct
4321 tgggcatcgg cactgtcctt gaccaagcag agactgcggg ggcgagactg gttgtgctcg
4381 ccactgctac ccctccgggc tccgtcactg tgtcccatcc taacatcgag gaggttgctc
4441 tgtccaccac cggagagatc cccttttacg gcaaggctat ccccctcgag gtgatcaagg
4501 ggggaagaca tctcatcttc tgccactcaa agaagaagtg cgacgagctc gccgcgaagc
4561 tggtcgcatt gggcatcaat gccgtggcct actaccgcgg tcttgacgtg tctgtcatcc
4621 cgaccagcgg cgatgttgtc gtcgtgtcga ccgatgctct catgactggc tttaccggcg
4681 acttcgactc tgtgatagac tgcaacacgt gtgtcactca gacagtcgat ttcagccttg
4741 accctacctt taccattgag acaaccacgc tcccccagga tgctgtctcc aggactcaac
4801 gccggggcag gactggcagg gggaagccag gcatctatag atttgtggca ccgggggagc
4861 gcccctccgg catgttcgac tcgtccgtcc tctgtgagtg ctatgacgcg ggctgtgctt
4921 ggtatgagct cacgcccgcc gagactacag ttaggctacg agcgtacatg aacacccccg
4981 ggcttcccgt gtgccaggac catcttgaat tttgggaggg cgtctttacg ggcctcactc
5041 atatagatgc ccacttttta tcccagacaa agcagagtgg ggagaacttt ccttacctgg
5101 tagcgtacca agccaccgtg tgcgctaggg ctcaagcccc tcccccatcg tgggaccaga
5161 tgtggaagtg tttgatccgc cttaaaccca ccctccatgg gccaacaccc ctgctataca
5221 gactgggcgc tgttcagaat gaagtcaccc tgacgcaccc aatcaccaaa tacatcatga
5281 catgcatgtc ggccgacctg gaggtcgtca cgagcacctg ggtgctcgtt ggcggcgtcc
5341 tggctgctct ggccgcgtat tgcctgtcaa caggctgcgt ggtcatagtg ggcaggatcg
5401 tcttgtccgg gaagccggca attatacctg acagggaggt tctctaccag gagttcgatg
5461 agatggaaga gtgctctcag cacttaccgt acatcgagca agggatgatg ctcgctgagc
5521 agttcaagca gaaggccctc ggcctcctgc agaccgcgtc ccgccatgca gaggttatca
5581 cccctgctgt ccagaccaac tggcagaaac tcgaggtctt ttgggcgaag cacatgtgga
5641 atttcatcag tgggatacaa tacttggcgg gcctgtcaac gctgcctggt aaccccgcca
5701 ttgcttcatt gatggctttt acagctgccg tcaccagccc actaaccact ggccaaaccc
5761 tcctcttcaa catattgggg gggtgggtgg ctgcccagct cgccgccccc ggtgccgcta
5821 ctgcctttgt gggtgctggc ctagctggcg ccgccatcgg cagcgttgga ctggggaagg
5881 tcctcgtgga cattcttgca gggtatggcg cgggcgtggc gggagctctt gtagcattca
5941 agatcatgag cggtgaggtc ccctccacgg aggacctggt caatctgctg cccgccatcc
6001 tctcgcctgg agcccttgta gtcggtgtgg tctgcgcagc aatactgcgc cggcacgttg
6061 gcccgggcga gggggcagtg caatggatga accggctaat agccttcgcc tcccggggga
6121 accatgtttc ccccacgcac tacgtgccgg agagcgatgc agccgcccgc gtcactgcca
6181 tactcagcag cctcactgta acccagctcc tgaggcgact gcatcagtgg ataagctcgg
6241 agtgtaccac tccatgctcc ggttcctggc taagggacat ctgggactgg atatgcgagg
6301 tgctgagcga ctttaagacc tggctgaaag ccaagctcat gccacaactg cctgggattc
6361 cctttgtgtc ctgccagcgc gggtataggg gggtctggcg aggagacggc attatgcaca
6421 ctcgctgcca ctgtggagct gagatcactg gacatgtcaa aaacgggacg atgaggatcg
6481 tcggtcctag gacctgcagg aacatgtgga gtgggacgtt ccccattaac gcctacacca
```

Fig. 13C

```
6541 cgggcccctg tactcccctt cctgcgccga actataagtt cgcgctgtgg agggtgtctg
6601 cagaggaata cgtggagata aggcgggtgg gggacttcca ctacgtatcg ggtatgacta
6661 ctgacaatct taaatgcccg tgccagatcc catcgcccga atttttcaca gaattggacg
6721 gggtgcgcct acacaggttt gcgccccctt gcaagccctt gctgcgggag gaggtatcat
6781 tcagagtagg actccacgag tacccggtgg ggtcgcaatt accttgcgag cccgaaccgg
6841 acgtagccgt gttgacgtcc atgctcactg atccctccca tataacagca gaggcggccg
6901 ggagaaggtt ggcgagaggg tcaccccctt ctatggccag ctcctcggct agccagctgt
6961 ccgctccatc tctcaaggca acttgcaccg ccaaccatga ctcccctgac gccgagctca
7021 tagaggctaa cctcctgtgg aggcaggaga tgggcggcaa catcaccagg gttgagtcag
7081 agaacaaagt ggtgattctg gactccttcg atccgcttgt ggcagaggag gatgagcggg
7141 aggtctccgt acctgcagaa attctgcgga agtctcggag attcgcccgg gccctgcccg
7201 tctgggcgcg gccggactac aacccccgc tagtagagac gtggaaaaag cctgactacg
7261 aaccacctgt ggtccatggc tgcccgctac cacctccacg gtcccctcct gtgcctccgc
7321 ctcggaaaaa gcgtacggtg gtcctcaccg aatcaaccct atctactgcc ttggccgagc
7381 ttgccaccaa aagttttggc agctcctcaa cttccggcat tacgggcgac aatacgacaa
7441 catcctctga gcccgcccct tctggctgcc cccccgactc cgacgttgag tcctattctt
7501 ccatgccccc cctggagggg gagcctgggg atccggatct cagcgacggg tcatggtcga
7561 cggtcagtag tggggccgac acggaagatg tcgtgtgctg ctcaatgtct tattcctgga
7621 caggcgcact cgtcaccccg tgcgctgcgg aagaacaaaa actgcccatc aacgcactga
7681 gcaactcgtt gctacgccat cacaatctgg tgtattccac cacttcacgc agtgcttgcc
7741 aaaggcagaa gaaagtcaca tttgacagac tgcaagttct ggacagccat taccaggacg
7801 tgctcaagga ggtcaaagca gcggcgtcaa aagtgaaggc taacttgcta tccgtagagg
7861 aagcttgcag cctgacgccc ccacattcag ccaaatccaa gtttggctat ggggcaaaag
7921 acgtccgttg ccatgccaga aaggccgtag cccacatcaa ctccgtgtgg aaagaccttc
7981 tggaagacag tgtaacacca atagacacta ccatcatggc caagaacgag gttttctgcg
8041 ttcagcctga gaaggggggt cgtaagccag ctcgtctcat cgtgttcccc gacctgggcg
8101 tgcgcgtgtg cgagaagatg gccctgtacg acgtggttag caagctcccc ctggccgtga
8161 tgggaagctc ctacggattc caatactcac caggacagcg ggttgaattc ctcgtgcaag
8221 cgtggaagtc caagaagacc ccgatggggt tctcgtatga tacccgctgt tttgactcca
8281 cagtcactga gagcgacatc cgtacggagg aggcaattta ccaatgttgt gacctggacc
8341 cccaagcccg cgtggccatc aagtccctca ctgagaggct ttatgttggg ggccctctta
8401 ccaattcaag gggggaaaac tgcggctacc gcaggtgccg cgcgagcggc gtactgacaa
8461 ctagctgtgg taacaccctc acttgctaca tcaaggcccg ggcagcctgt cgagccgcag
8521 ggctccagga ctgcaccatg ctcgtgtgtg gcgacgactt agtcgttatc tgtgaaagtg
8581 cgggggtcca ggaggacgcg gcgagcctga gagccttcac ggaggctatg accaggtact
8641 ccgccccccc cggggacccc ccacaaccag aatacgactt ggagcttata acatcatgct
8701 cctccaacgt gtcagtcgcc cacgacggcg ctggaaagag ggtctactac cttacccgtg
8761 accctacaac cccccctcgcg agagccgcgt gggagacagc aagacacact ccagtcaatt
8821 cctggctagg caacataatc atgtttgccc ccacactgtg ggcgaggatg atactgatga
8881 cccatttctt tagcgtcctc atagccaggg atcagcttga acaggctctt aactgtgaga
8941 tctacggagc ctgctactcc atagaaccac tggatctacc tccaatcatt caaagactcc
9001 atggcctcag cgcattttca ctccacagtt actctccagg tgaaatcaat agggtggccg
9061 catgcctcag aaaacttggg gtcccgccct tgcgagcttg gagacaccgg gcccggagcg
9121 tccgcgctag gcttctgtcc agaggaggca gggctgccat atgtggcaag tacctcttca
9181 actgggcagt aagaacaaag ctcaaactca ctccaatagc ggccgctggc cggctggact
9241 tgtccggttg gttcacggct ggctacagcg gggagacat ttatcacagc gtgtctcatg
9301 cccggccccg ctggttctgg ttttgcctac tcctgctcgc tgcaggggta ggcatctacc
9361 tcctccccaa ccgatgaagg ttggggtaaa cactccggcc tcttaagcca tttcctgttt
9421 ttttttttt tttttttt tttttctttt ttttttctt tcctttcctt ctttttttcc
9481 tttcttttc ccttctttaa tggtggctcc atcttagccc tagtcacggc tagctgtgaa
9541 aggtccgtga gccgcatgac tgcagagagt gctgatactg gcctctctgc agatcatgt
```

Fig. 13D

MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRL
GVRATRKTSERSQPRGRRQPIPKARRPEGRTWAQPGYPWPLYGNEGCGWAGWLLSPRG
SRPSWGPTDPRRRSRNLGKVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLED
GVNYATGNLPGCSFSIFLLALLSCLTVPASAYQVRNSSGLYHVTNDCPNSSIVYEAAD
AILHTPGCVPCVREGNASRCWVAVTPTVATRDGKLPTTQLRRHIDLLVGSATLCSALY
VGDLCGSVFLVGQLFTFSPRRHWTTQDCNCSIYPGHITGHRMAWDMMMNWSPTAALVV
AQLLRIPQAIMDMIAGAHWGVLAGIAYFSMVGNWAKVLVVLLLFAGVDAETHVTGGNA
GRTTAGLVGLLTPGAKQNIQLINTNGSWHINSTALNCNESLNTGWLAGLFYQHKFNSS
GCPERLASCRRLTDFAQGWGPISYANGSGLDERPYCWHYPPRPCGIVPAKSVCGPVYC
FTPSPVVVGTTDRSGAPTYSWGANDTDVFVLNNTRPPLGNWFGCTWMNSTGFTKVCGA
PPCVIGGVGNNTLLCPTDCFRKHPEATYSRCGSGPWITPRCMVDYPYRLWHYPCTINY
TIFKVRMYVGGVEHRLEAACNWTRGERCDLEDRDRSELSPLLLSTTQWQVLPCSFTTL
PALSTGLIHLHQNIVDVQYLYGVGSSIASWAIKWEYVVLLFLLLADARVCSCLWMMLL
ISQAEAALENLVILNAASLAGTHGLVSFLVFFCFAWYLKGRWVPGAVYALYGMWPLLL
LLLALPQRAYALDTEVAASCGGVVLVGLMALTLSPYYKRYISWCMWWLQYFLTRVEAQ
LHVWVPPLNVRGGRDAVILLMCVVHPTLVFDITKLLLAIFGPLWILQASLLKVPYFVR
VQGLLRICALARKIAGGHYVQMAIIKLGALTGTYVYNHLTPLRDWAHNGLRDLAVAVE
PVVFSRMETKLITWGADTAACGDIINGLPVSARRGQEILLGPADGMVSKGWRLLAPIT
AYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTATQTFLATCINGVCWTVYHGAGTR
TIASPKGPVIQMYTNVDQDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRRG
DSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGLFRAAVCTRGVAKAVDFIPVENLGTT
MRSPVFTDNSSPPAVPQSFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLNPSVAATL
GFGAYMSKAHGVDPNIRTGVRTITTGSPITYSTYGKFLADGGCSGGAYDIIICDECHS
TDATSILGIGTVLDQAETAGARLVVLATATPPGSVTVSHPNIEEVALSTTGEIPFYGK
AIPLEVIKGGRHLIFCHSKKKCDELAAKLVALGINAVAYYRGLDVSVIPTSGDVVVVS
TDALMTGFTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTTLPQDAVSRTQRRGRTGRG
KPGIYRFVAPGERPSGMFDSSVLCECYDAGCAWYELTPAETTVRLRAYMNTPGLPVCQ
DHLEFWEGVFTGLTHIDAHFLSQTKQSGENFPYLVAYQATVCARAQAPPPSWDQMWKC
LIRLKPTLHGPTPLLYRLGAVQNEVTLTHPITKYIMTCMSADLEVVTSTWVLVGGVLA

Fig. 13E

ALAAYCLSTGCVVIVGRIVLSGKPAIIPDREVLYQEFDEMEECSQHLPYIEQGMMLAE
QFKQKALGLLQTASRHAEVITPAVQTNWQKLEVFWAKHMWNFISGIQYLAGLSTLPGN
PAIASLMAFTAAVTSPLTTGQTLLFNILGGWVAAQLAAPGAATAFVGAGLAGAAIGSV
GLGKVLVDILAGYGAGVAGALVAFKIMSGEVPSTEDLVNLLPAILSPGALVVGVVCAA
ILRRHVGPGEGAVQWMNRLIAFASRGNHVSPTHYVPESDAAARVTAILSSLTVTQLLR
RLHQWISSECTTPCSGSWLRDIWDWICEVLSDFKTWLKAKLMPQLPGIPFVSCQRGYR
GVWRGDGIMHTRCHCGAEITGHVKNGTMRIVGPRTCRNMWSGTFPINAYTTGPCTPLP
APNYKFALWRVSAEEYVEIRRVGDFHYVSGMTTDNLKCPCQIPSPEFFTELDGVRLHR
FAPPCKPLLREEVSFRVGLHEYPVGSQLPCEPEPDVAVLTSMLTDPSHITAEAAGRRL
ARGSPPSMASSSASQLSAPSLKATCTANHDSPDAELIEANLLWRQEMGGNITRVESEN
KVVILDSFDPLVAEEDEREVSVPAEILRKSRRFARALPVWARPDYNPPLVETWKKPDY
EPPVVHGCPLPPPRSPPVPPPRKKRTVVLTESTLSTALAELATKSFGSSSTSGITGDN
TTTSSEPAPSGCPPDSDVESYSSMPPLEGEPGDPDLSDGSWSTVSSGADTEDVVCCSM
SYSWTGALVTPCAAEEQKLPINALSNSLLRHHNLVYSTTSRSACQRQKKVTFDRLQVL
DSHYQDVLKEVKAAASKVKANLLSVEEACSLTPPHSAKSKFGYGAKDVRCHARKAVAH
INSVWKDLLEDSVTPIDTTIMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALY
DVVSKLPLAVMGSSYGFQYSPGQRVEFLVQAWKSKKTPMGFSYDTRCFDSTVTESDIR
TEEAIYQCCDLDPQARVAIKSLTERLYVGGPLTNSRGENCGYRRCRASGVLTTSCGNT
LTCYIKARAACRAAGLQDCTMLVCGDDLVVICESAGVQEDAASLRAFTEAMTRYSAPP
GDPPQPEYDLELITSCSSNVSVAHDGAGKRVYYLTRDPTTPLARAAWETARHTPVNSW
LGNIIMFAPTLWARMILMTHFFSVLIARDQLEQALNCEIYGACYSIEPLDLPPIIQRL
HGLSAFSLHSYSPGEINRVAACLRKLGVPPLRAWRHRARSVRARLLSRGGRAAICGKY
LFNWAVRTKLKLTPIAAAGRLDLSGWFTAGYSGGDIYHSVSHARPRWFWFCLLLLAAG
VGIYLLPNR"

Fig. 14A

```
   1 gccagccccc tgatgggggc gacactccac catgaatcac tcccctgtga ggaactactg
  61 tcttcacgca gaaagcgtct agccatggcg ttagtatgag tgtcgtgcag cctccaggac
 121 cccccctccc gggagagcca tagtggtctg cggaaccggt gagtacaccg gaattgccag
 181 gacgaccggg tcctttcttg gataaacccg ctcaatgcct ggagatttgg gcgtgccccc
 241 gcaagactgc tagccgagta gtgttgggtc gcgaaaggcc ttgtggtact gcctgatagg
 301 gtgcttgcga gtgccccggg aggtctcgta gaccgtgcac catgagcacg aatcctaaac
 361 ctcaaagaaa aaccaaacgt aacaccaacc gtcgcccaca ggacgtcaag ttcccgggtg
 421 gcggtcagat cgttggtgga gtttacttgt tgccgcgcag gggccctaga ttgggtgtgc
 481 gcgcgacgag gaagacttcc gagcggtcgc aacctcgagg tagacgtcag cctatcccca
 541 aggcacgtcg gcccgagggc aggacctggg ctcagcccgg gtacccttgg cccctctatg
 601 gcaatgaggg ttgcgggtgg gcgggatggc tcctgtctcc ccgtggctct cggcctagct
 661 ggggccccac agacccccgg cgtaggtcgc gcaatttggg taaggtcatc gataccctta
 721 cgtgcggctt cgccgacctc atggggtaca taccgctcgt cggcgcccct cttggaggcg
 781 ctgccagggc cctggcgcat ggcgtccggg ttctggaaga cggcgtgaac tatgcaacag
 841 ggaaccttcc tggttgctct ttctctatct tccttctggc cctgctctct tgcctgactg
 901 tgcccgcttc agcctaccaa gtgcgcaatt cctcggggct ttaccatgtc accaatgatt
 961 gccctaactc gagtgttgtg tacgaggcgg ccgatgccat cctgcacact ccggggtgtg
1021 tcccttgcgt tcgcgagggt aacgcctcga ggtgttgggt ggcggtgacc cccacggtgg
1081 ccaccaggga cggcaaactc cccacaacgc agcttcgacg tcatatcgat ctgcttgtcg
1141 ggagcgccac cctctgctcg gccctctacg tgggggacct gtgcgggtct gtctttcttg
1201 ttggtcaact gtttaccttc tctcccaggc accactggac gacgcaagac tgcaattgtt
1261 ctatctatcc cggccatata acgggtcatc gcatggcatg gaatatgatg atgaactggt
1321 cccctacggc agcgttggtg gtagctcagc tgctccgaat cccacaagcc atcatggaca
1381 tgatcgctgg cgcccactgg ggagtcctgg cgggcataaa gtatttctcc atggtgggga
1441 actgggcgaa ggtcctggta gtgctgctgc tatttgccgg cgtcgacgcg gaaacccacg
1501 tcaccggggg aaatgccggc cgcaccacgg ctgggcttgt tggtctcctt acaccaggcg
1561 ccaagcagaa catccaactg atcaacacca acggcagttg gcacatcaat agcacggcct
1621 tgaactgcaa tgaaagcctt aacaccggct ggttagcagg gctcttctat cagcacaaat
1681 tcaactcttc aggctgtcct gagaggttgg ccagctgccg acgccttacc gattttgccc
1741 agggctgggg tcctatcagt tatgccaacg gaagcggcct cgacgaacgc ccctactgct
1801 ggcactaccc tccaagacct tgtggcattg tgcccgcaaa gagcgtgtgt ggcccggtat
1861 attgcttcac tcccagcccc gtggtggtgg gaacgaccga caggtcgggc gcgcctacct
1921 acagctgggg tgcaaatgat acggatgtct tcgtccttaa caacaccagg ccaccgctgg
1981 gcaattggtt cggttgtacc tggatgaact caactggatt caccaaagtg tgcggagcgc
2041 cccttgtgt  catcggaggg gtgggcaaca acaccttgct ctgccccact gattgcttcc
2101 gcaaatatcc ggaagccaca tactctcggt gcggctccgg tcccaggatt acacccaggt
2161 gcatggtcga ctacccgtat aggctttggc actatccttg taccatcaat tacaccatat
2221 tcaaagtcag gatgtacgtg ggaggggtcg agcacaggct ggaagcggcc tgcaactgga
2281 cgcggggcga acgctgtgat ctggaagaca gggacaggtc cgagctcagc ccgttgctgc
2341 tgtccaccac acagtggcag gtccttccgt gttctttcac gaccctgcca gccttgtcca
2401 ccggcctcat ccacctccac cagaacattg tggacgtgca gtacttgtac ggggtagggt
2461 caagcatcgc gtcctgggcc attaagtggg agtacgtcgt tctcctgttc cttctgcttg
2521 cagacgcgcg cgtctgttcc tgcttgtgga tgatgttact catatcccaa gcggaggcgg
2581 ctttggagaa cctcgtaata ctcaatgcag catccctggc cgggacgcat ggtcttgtgt
2641 ccttcctcgt gttcttctgc tttgcgtggt atctgaaggg taggtgggtg cccggagcgg
2701 tctacgccct ctacgggatg tggcctctcc tcctgctcct gctggcgttg cctcagcggg
2761 catacgcact ggacacggag gtggccgcgt cgtgtggcgg cgttgttctt gtcgggttaa
2821 tggcgctgac tctgtcgcca tattacaagc gctatatcag ctggtgcatg tggtggcttc
2881 agtattttct gaccagagta gaagcgcaac tgcacgtgtg ggttcccccc ctcaacgtcc
2941 gggggggcg  cgatgccgtc atcttactca cgtgtgtagt acacccggcc ctggtatttg
3001 acatcaccaa actactcctg gccatcttcg gacccctttg gattcttcaa gccagtttgc
3061 ttaaagtccc ctacttcgtg cgcgttcaag gccttctccg gatctgcgcg ctagcgcgga
```

Fig. 14B

```
3121 agatagccgg aggtcattac gtgcaaatgg ccatcatcaa gttaggggcg cttactggca
3181 cctgtgtgta taaccatctc gctcctcttc gagactgggc gcacaacggc ctgcgagatc
3241 tggccgtggc tgtggaacca gtcgtcttct cccgaatgga gaccaagctc atcacgtggg
3301 gggcagatac cgccgcgtgc ggtgacatca tcaacggctt gcccgtctct gcccgtaggg
3361 gccaggagat actgcttggg ccagccgacg gaatggtctc caaggggtgg aggttgctgg
3421 cgcccatcac ggcgtacgcc cagcagacga gaggcctcct agggtgtata atcaccagcc
3481 tgactggccg ggacaaaaac caagtggagg gtgaggtcca gatcgtgtca actgctaccc
3541 agaccttcct ggcaacgtgc atcaatgggg tatgctggac tgtctaccac ggggccggaa
3601 cgaggaccat cgcatcaccc aagggtcctg tcatccagac gtataccaat gtggatcaag
3661 acctcgtggg ctggcccgct cctcaaggtt cccgctcatt gacaccctgc acctgcggct
3721 cctcggacct ttacctggtc acgaggcacg ccgatgtcat tcccgtgcgc cggcgaggtg
3781 atagcagggg tagcctgctt tcgccccggc ccatttccta cttgaaaggc tcctcggggg
3841 gtccgctgtt gtgccccacg ggacacgccg tgggcctatt cagggccgcg gtgtgcaccc
3901 gtggagtggc taaggcggtg gactttatcc ctgtggagaa cctagagaca accatgagat
3961 ccccggtgtt cacggacaac tcctctccac cagcagtgcc ccagagcttc caggtggccc
4021 acctgcatgc tcccaccggc agcggtaaga gcaccaaggt cccggctgcg tacgcagcca
4081 agggctacaa ggtgttggtg ctcaacccct ctgttgctgc aacactgggc tttggtgctt
4141 acatgtccaa ggcccatggg gttgatccta atatcaggac cggggtgaga acaattacca
4201 ctggcagccc catcacgtac tccacctacg gcaagttcct tgccgacgcc gggtgctcag
4261 gaggtgctta tgacataata atttgtgacg agtgccactc cacggatgcc acatccatct
4321 cgggcatcgg cactgtcctt gaccaagcag agactgcggg ggcgagactg gttgtgctcg
4381 ccactgctac ccctccgggc tccgtcactg tgtcccatcc taacatcgag gaggttgctc
4441 tgtccaccac cggagagatc ccttttacg gcaaggctat cccccctcgag gtgatcaagg
4501 ggggaagaca tctcatcttc tgccactcaa agaagaagtg cgacgagctc gccgcgaagc
4561 tggtcgcatt gggcatcaat gccgtggcct actaccgcgg tcttgacgtg tctgtcatcc
4621 cgaccagcgg cgatgttgtc gtcgtgtcga ccgatgctct catgactggc tttaccggcg
4681 acttcgactc tgtgatagac tgcaacacgt gtgtcactca gacagtcgat tttagccttg
4741 accctacctt taccattgag acaaccacgc tcccccagga tgctgtctcc aggactcaac
4801 gccgggggcag gactggcagg gggaagccag gcatctatag atttgtggca ccgggggagc
4861 gcccctccgg catgttcgac tcgtccgtcc tctgtgagtg ctatgacgcg ggctgtgctt
4921 ggtatgagct cacgcccgcc gagactacag ttaggctacg agcgtacatg aacaccccgg
4981 ggcttcccgt gtgccaggac catcttggat tttgggaggg cgtctttacg ggcctcactc
5041 atatagatgc ccactttcta tcccagacaa agcagagtgg ggagaacttt ccttacctgg
5101 tagcgtacca agccaccgtg tgcgctaggg ctcaagcccc tcccccatcg tgggaccaga
5161 tgcggaagtg tttgatccgc cttaaaccca ccctccatgg gccaacaccc ctgctataca
5221 gactgggcgc tgttcagaat gaagtcaccc tgacgcaccc aatcaccaaa tacatcatga
5281 catgcatgtc ggccgacctg gaggtcgtca cgagcacctg ggtgctcgtt ggcggcgtcc
5341 tggctgctct ggccgcgtat tgcctgtcaa caggctgcgt ggtcatagtg ggcaggatcg
5401 tcttgtccgg gaagccggca attatacctg acagggaggt tctctaccag gagttcgatg
5461 agatggaaga gtgctctcag cacttaccgt acatcgagca agggatgatg ctcgctgagc
5521 agttcaagca gaaggccctc ggcctcctgc agaccgcgtc ccgccatgca gaggttatca
5581 cccctgctgt ccagaccaac tggcagaaac tcgaggtctt ttgggcgaag cacatgtgga
5641 atttcatcag tgggatacaa tacttggcgg gcctgtcaac gctgcctggt aaccccgcca
5701 ttgcttcatt gatggctttt acagctgccg tcaccagccc actaaccact ggccaaaccc
5761 tcctcttcaa catattgggg gggtgggtgg ctgcccagct cgccgccccc ggtgccgcta
5821 ccgcctttgt gggcgctggc ttagctggcg ccgcactcga cagcgttgga ctggggaagg
5881 tcctcgtgga cattcttgca ggctatggcg cgggcgtggc gggagctctt gtggcattca
5941 agatcatgag cggtgaggtc ccctccacgg aggacctggt caatctgctg cccgccatcc
6001 tctcacctgg agcccttgca gtcggtgtgg tctttgcatc aatactgcgc cggcgtgttg
6061 gcccgggcga gggggcagtg caatggatga accggctaat agccttcgcc tcccgggggga
6121 accatgtttc cccccacacac tacgtgccgg agagcgatgc agccgcccgc gtcactgcca
6181 tactcagcag cctcactgta acccagctcc tgaggcgact gcatcagtgg ataagctcgg
6241 agtgtaccac tccatgctcc ggttcctggc taagggacat ctgggactgg atatgcgagg
6301 tgctgagcga ctttaagacc tggctgaaag ccaagctcat gccacaactg cctgggattc
6361 cctttgtgtc ctgccagcgc gggtataggg gggtctggcg aggagacggc attatgcaca
6421 ctcgctgcca ctgtggagct gagatcactg gacatgtcaa aaacgggacg atgaggatcg
6481 tcggtcctag gacctgcaag aacatgtgga gtgggacgtt cttcattaat gcctacacca
```

Fig. 14C

6541 cgggcccctg tactcccctt cctgcgccga actataagtt cgcgctgtgg agggtgtctg
6601 cagaggaata cgtggagata aggcgggtgg gggacttcca ctacgtatcg ggcatgacta
6661 ctgacaatct caaatgcccg tgccagatcc catcgcccga atttttcaca gaattggacg
6721 gggtgcgcct acataggttt gcgcccccctt gcaagccctt gctgcgggag gaggtatcat
6781 tcagagtagg actccacgag tacccggtgg ggtcgcaatt accttgcgag cccgaaccgg
6841 acgtagccgt gttgacgtcc atgctcactg atccctccca tataacagca gaggcggccg
6901 ggagaaggtt ggcgagaggg tcaccccctt ctatggccag ctcctcggct agccagctgt
6961 ccgctccatc tctcaaggca acttgcaccg ccaaccatga ctcccctgac gccgagctca
7021 tagaggctaa cctcctgtgg aggcaggaga tgggcggcaa catcaccagg gttgagtcag
7081 agaacaaagt ggtgattctg gactccttcg atccgcttgt ggcagaggag gatgagcggg
7141 aggtctccgt acccgcagaa attctgcgga agtctcggag attcgcccca gccctgcccg
7201 tctgggcgcg gccggactac aacccccctgc tagtagagac gtggaaaaag cctgactacg
7261 aaccacctgt ggtccatggc tgcccgctac cacctccacg gtcccctcct gtgcctccgc
7321 ctcggaaaaa gcgtacggtg gtcctcaccg aatcaaccct acctactgcc ttggccgagc
7381 ttgccaccaa aagttttggc agctcctcaa cttccggcat tacgggcgac aatacgacaa
7441 catcctctga gcccgcccct tctggctgcc cccccgactc cgacgttgag tcctattctt
7501 ccatgccccc cctggagggg gagcctgggg atccggatct cagcgacggg tcatggtcga
7561 cggtcagtag tggggccgac acggaagatg tcgtgtgctg ctcaatgtct tattcctgga
7621 caggcgcact cgtcaccccg tgcgctgcgg aggaacaaaa actgcccatc aacgcactga
7681 gcaactcgtt gctacgccat cacaatctgg tgtattccac cacttcacgc agtgcttgcc
7741 aaaggaagaa gaaagtcaca tttgacagac tgcaagttct ggacagccat taccaggacg
7801 tgctcaagga ggtcaaagca gcggcgtcaa aagtgaaggc taacttgcta tccgtagagg
7861 aagcttgcag cctggcgccc ccacattcag ccaaatccaa gtttggctat ggggcaaaag
7921 acgtccgttg ccatgccaga aaggccgtag cccacatcaa ctccgtgtgg aaagaccttc
7981 tggaagacag tgtaacacca atagacacta ccatcatggc caagaacgag gttttctgcg
8041 ttcagcctga gaagggggt cgtaagccag ctcgtctcat cgtgttcccc gacctgggcg
8101 tgcgcgtgtg cgagaagatg gccctgtacg acgtggttag caagctcccc ttggccgtga
8161 tgggaagctc ctacggattc caatactcac caggacagcg ggttgaattc ctcgtgcaag
8221 cgtggaagtc caagaagacc ccgatggggc tctcgtatga tacccgctgt tttgactcca
8281 cagtcactga gagcgacatc cgtacggagg aggcaattta ccaatgttgt gacctggacc
8341 cccaagcccg cgtggccatc aagtccctca ctgagaggct ttatgttggg ggccctctta
8401 ctaattcaag gggggaaaac tgcggctacc gcaggtgccg cgcgagcaga gtactgacaa
8461 ctagctgtgg taacaccctc actcgctaca tcaaggcccg ggcagcctgt cgagccgcag
8521 ggctccagga ctgcaccatg ctcgtgtgtg gcgacgactt agtcgttatc tgtgaaagtg
8581 cggggtcca ggaggacgcg gcgagcctga gagccttcac ggaggctatg accaggtact
8641 ccgccccccc cggggacccc ccacaaccag aatacgactt ggagcttata acatcatgct
8701 cctccaacgt gtcagtcgcc cacgacggcg ctggaaagag ggtctactac cttacccgtg
8761 accctacaac ccccctcgcg agagccgcgt gggagacagc aagacacact ccagtcaatt
8821 cctggctagg caacataatc atgtttgccc ccacactgtg ggcgaggatg atactgatga
8881 cccacttctt tagcgtcctc atagccaggg atcagcttga acaggctctc aactgcgaga
8941 tctacggagc ctgctactcc atagaaccac tggatctacc tccaatcatt caaagactcc
9001 atggcctcag cgcattttca ctccacagtt actctccagg tgaaattaat agggtggccg
9061 catgcctcag aaaacttggg gtcccgccct tgcgagcttg gagacaccgg gcctggagcg
9121 tccgcgctag gcttctggcc agaggaggca aggctgccat atgtggcaag tacctcttca
9181 actgggcagt aagaacaaag ctcaaactca ctccgataac ggccgctggc cggctggact
9241 tgtccggctg gttcacggct ggctacagcg gggagacat ttatcacagc gtgtctcatg
9301 cccggccccg ctggttctgg ttttgcctac tcctgcttgc tgcaggggta ggcatctacc
9361 tcctccccaa ccgatgaaga ttgggctaac cactccaggc caataggcca ttccct

Fig. 14D

MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRL
GVRATRKTSERSQPRGRRQPIPKARRPEGRTWAQPGYPWPLYGNEGCGWAGWLLSPRG
SRPSWGPTDPRRRSRNLGKVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLED
GVNYATGNLPGCSFSIFLLALLSCLTVPASAYQVRNSSGLYHVTNDCPNSSVVYEAAD
AILHTPGCVPCVREGNASRCWVAVTPTVATRDGKLPTTQLRRHIDLLVGSATLCSALY
VGDLCGSVFLVGQLFTFSPRHHWTTQDCNCSIYPGHITGHRMAWNMMMNWSPTAALVV
AQLLRIPQAIMDMIAGAHWGVLAGIKYFSMVGNWAKVLVVLLLFAGVDAETHVTGGNA
GRTTAGLVGLLTPGAKQNIQLINTNGSWHINSTALNCNESLNTGWLAGLFYQHKFNSS
GCPERLASCRRLTDFAQGWGPISYANGSGLDERPYCWHYPPRPCGIVPAKSVCGPVYC
FTPSPVVVGTTDRSGAPTYSWGANDTDVFVLNNTRPPLGNWFGCTWMNSTGFTKVCGA
PPCVIGGVGNNTLLCPTDCFRKYPEATYSRCGSGPRITPRCMVDYPYRLWHYPCTINY
TIFKVRMYVGGVEHRLEAACNWTRGERCDLEDRDRSELSPLLLSTTQWQVLPCSFTTL
PALSTGLIHLHQNIVDVQYLYGVGSSIASWAIKWEYVVLLFLLLADARVCSCLWMMLL
ISQAEAALENLVILNAASLAGTHGLVSFLVFFCFAWYLKGRWVPGAVYALYGMWPLLL
LLLALPQRAYALDTEVAASCGGVVLVGLMALTLSPYYKRYISWCMWWLQYFLTRVEAQ
LHVWVPPLNVRGGRDAVILLTCVVHPALVFDITKLLLAIFGPLWILQASLLKVPYFVR
VQGLLRICALARKIAGGHYVQMAIIKLGALTGTCVYNHLAPLRDWAHNGLRDLAVAVE
PVVFSRMETKLITWGADTAACGDIINGLPVSARRGQEILLGPADGMVSKGWRLLAPIT
AYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTATQTFLATCINGVCWTVYHGAGTR
TIASPKGPVIQTYTNVDQDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRRG
DSRGSLLSPRPISYLKGSSGGPLLCPTGHAVGLFRAAVCTRGVAKAVDFIPVENLETT
MRSPVFTDNSSPPAVPQSFQVAHLHAPTGSGKSTKVPAAYAAKGYKVLVLNPSVAATL
GFGAYMSKAHGVDPNIRTGVRTITTGSPITYSTYGKFLADAGCSGGAYDIIICDECHS
TDATSISGIGTVLDQAETAGARLVVLATATPPGSVTVSHPNIEEVALSTTGEIPFYGK
AIPLEVIKGGRHLIFCHSKKKCDELAAKLVALGINAVAYYRGLDVSVIPTSGDVVVVS
TDALMTGFTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTTLPQDAVSRTQRRGRTGRG
KPGIYRFVAPGERPSGMFDSSVLCECYDAGCAWYELTPAETTVRLRAYMNTPGLPVCQ
DHLGFWEGVFTGLTHIDAHFLSQTKQSGENFPYLVAYQATVCARAQAPPPSWDQMRKC
LIRLKPTLHGPTPLLYRLGAVQNEVTLTHPITKYIMTCMSADLEVVTSTWVLVGGVLA

Fig. 14E

ALAAYCLSTGCVVIVGRIVLSGKPAIIPDREVLYQEFDEMEECSQHLPYIEQGMMLAE
QFKQKALGLLQTASRHAEVITPAVQTNWQKLEVFWAKHMWNFISGIQYLAGLSTLPGN
PAIASLMAFTAAVTSPLTTGQTLLFNILGGWVAAQLAAPGAATAFVGAGLAGAALDSV
GLGKVLVDILAGYGAGVAGALVAFKIMSGEVPSTEDLVNLLPAILSPGALAVGVVFAS
ILRRRVGPGEGAVQWMNRLIAFASRGNHVSPTHYVPESDAAARVTAILSSLTVTQLLR
RLHQWISSECTTPCSGSWLRDIWDWICEVLSDFKTWLKAKLMPQLPGIPFVSCQRGYR
GVWRGDGIMHTRCHCGAEITGHVKNGTMRIVGPRTCKNMWSGTFFINAYTTGPCTPLP
APNYKFALWRVSAEEYVEIRRVGDFHYVSGMTTDNLKCPCQIPSPEFFTELDGVRLHR
FAPPCKPLLREEVSFRVGLHEYPVGSQLPCEPEPDVAVLTSMLTDPSHITAEAAGRRL
ARGSPPSMASSSASQLSAPSLKATCTANHDSPDAELIEANLLWRQEMGGNITRVESEN
KVVILDSFDPLVAEEDEREVSVPAEILRKSRRFAPALPVWARPDYNPLLVETWKKPDY
EPPVVHGCPLPPPRSPPVPPPRKKRTVVLTESTLPTALAELATKSFGSSSTSGITGDN
TTTSSEPAPSGCPPDSDVESYSSMPPLEGEPGDPDLSDGSWSTVSSGADTEDVVCCSM
SYSWTGALVTPCAAEEQKLPINALSNSLLRHHNLVYSTTSRSACQRKKKVTFDRLQVL
DSHYQDVLKEVKAAASKVKANLLSVEEACSLAPPHSAKSKFGYGAKDVRCHARKAVAH
INSVWKDLLEDSVTPIDTTIMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALY
DVVSKLPLAVMGSSYGFQYSPGQRVEFLVQAWKSKKTPMGLSYDTRCFDSTVTESDIR
TEEAIYQCCDLDPQARVAIKSLTERLYVGGPLTNSRGENCGYRRCRASRVLTTSCGNT
LTRYIKARAACRAAGLQDCTMLVCGDDLVVICESAGVQEDAASLRAFTEAMTRYSAPP
GDPPQPEYDLELITSCSSNVSVAHDGAGKRVYYLTRDPTTPLARAAWETARHTPVNSW
LGNIIMFAPTLWARMILMTHFFSVLIARDQLEQALNCEIYGACYSIEPLDLPPIIQRL
HGLSAFSLHSYSPGEINRVAACLRKLGVPPLRAWRHRAWSVRARLLARGGKAAICGKY
LFNWAVRTKLKLTPITAAGRLDLSGWFTAGYSGGDIYHSVSHARPRWFWFCLLLLAAG
VGIYLLPNR"

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
16 June 2005 (16.06.2005)

PCT

(10) International Publication Number
WO 2005/053516 A3

(51) International Patent Classification[7]: C07H 21/02, 21/04, C12P 19/34, 21/00, C12N 15/09, 15/82, 15/85, 15/00, C12Q 1/70, 1/68

(21) International Application Number: PCT/US2004/040120

(22) International Filing Date: 1 December 2004 (01.12.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/525,989 1 December 2003 (01.12.2003) US

(71) Applicant (*for all designated States except US*): **BOARDS OF REGENTS, THE UNIVERSITY OF TEXAS SYSTEM** [US/US]; 201 West 7th Street, Austin, TX 78701 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): **LEMON, Stanley, M.** [US/US]; 1517 Bayou Shore Drive, Galveston, TX 77551 (US). **YI, MinKyung** [KR/US]; 7700 Seawall Blvd.#301, Galveston, TX 77551 (US).

(74) Agent: **PROVENCE, David, L.**; Mueting, Raasch & Gebhardt, P.A., P.O. Box 581415, Minneapolis, MN 55454-1415 (US).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) Date of publication of the international search report: 29 December 2005

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2005/053516 A3

(54) Title: REPLICATION COMPETENT HEPATITIS C VIRUS AND METHODS OF USE

(57) Abstract: The present invention provides replication competent polynucleotides that include a coding sequence encoding a hepatitis C virus polyprotein having adaptive mutations. The invention also includes methods for making replication competent polynucleotides, identifying a compound that inhibits replication of a replication competent polynucleotide, selecting a replication competent polynucleotide, and detecting a replication competent polynucleotide.

INTERNATIONAL SEARCH REPORT	International application No. PCT/US04/40120

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C07H 21/02, 21/04; C12P 19/34, 21/00; C12N 15/09, 15/82, 15/85, 15/00; C12Q 1/70, 1/68

US CL : 536/23.1, 23.72; 435440, 441, 442455, 91.1, 91.4, 91.4291.51, 69.1,69.270.1, 455, 5, 6, 94

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.1, 23.72; 435440. 441, 442455, 91.1, 91.4, 91.4291.51. 69.1,69.270.1, 455, 5, 6, 94

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	US6,689,559B2 (WIMMER et al) 10 February 2004 (10.02.2004), Fig. 2-4, Claims 1-12, columns 11-13).	1-36
X	US 20020098202A1 (WIMMER et al) 24 October 2002 (24.10.2002), see Fig. 1-4, claims 1-52)	1-36
A	IKEDA. M. et al. Selection Subgenomic and Genome-Length Dicsitronic RNAs Derived from an Infectiou Molecular Clone of the HCV-N Strain of hepatitis C virus Replicate Effeciently in Cultureed Huh7 Cells. J. Virol. March 2002, Vol. 76. No. 6, pages 2997-3006.	
A	LOHMANN. V. et al. Mutations in Hepatitis C Virus RNAs conferring Cell Culture Adaptation. J. Virol. February 2001, Vol. 75. No. 3, pages 1437-1449.	
X	BUKH, J. et al. Mutations that permit effecient replication of hepatitis C virus RNA in Huh-7 cells prevent productive replication in chimpazees, PNAS, October 2002, Voil. 99. No. 22, pages 14416-14421, see page 14416, 2nd column, last paragraphand 1st column, 2nd paragraph.	1-22, 7, 9-10, 14-17,
A	KRIEGER. N. et al. Enhancement of Hepatitis C Virus RNA Replication by Cell Culture-Adaptive Mutations. J. Virol. May 2001, Vol. 75. No. 10, pages 4614-4624.	

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier application or patent published on or after the international filing date		
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed	"&"	document member of the same patent family

Date of the actual completion of the international search 17 August 2005 (17.08.2005)	Date of mailing of the international search report 21 SEP 2005
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450 Facsimile No. (703) 305-3230	Authorized officer Bao Qun Li Telephone No. (571) 272-1600

Form PCT/ISA/210 (second sheet) (January 2004)

Continuation of B. FIELDS SEARCHED Item 3:
WEST, MEDLINE, CAPLUS
Search terms: HCV replicon, in vitro, synthesis, construct, expression system